PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

OCT 17 1938

VOL. XXIX SEPTEMBER, 1938 NO. 6

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

Language Acquisition by Deaf Children as Related to Hearing Loss and Age of Onset 401
NOEL KEYS AND LOUISE BOULWARE

Effect of Age on the Speed-power Relationship with Reference to Tests of Intelligence 413
RALPH R. BROWN

Meaning as a Factor in Learning 419
H. B. REED

Systems of Regression Equations for the Estimation of Factors . . . 431
HARRY H. HARMAN

Scale Calibration by the Thomson Method 442
KATHERINE PRESTON BRADWAY

The Concept of Intelligence 449
D. G. RYANS

A Comparative Study of the Psychoneurotic Responses of Deaf and Hearing Children 459
N. NORTON SPRINGER

A Preliminary Study of Relations between Attitude, Adjustment, and Vocational Interest Tests 467
JOHN G. DARLEY

Book Reviews 474

$6.00 per Year · Published Monthly September to May

WARWICK & YORK, INC.

BALTIMORE, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as Second Class Matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

Established 1910

EDITORS

JACK W. DUNLAP
University of Rochester

HAROLD E. JONES
University of California

PERCIVAL M. SYMONDS
Teachers College,
Columbia University

H. E. BUCHHOLZ, *Managing Editor*

THE price of the Journal is $6.00 a year in the United States; $6.40 in foreign countries. Part-year subscriptions are 90 cents for each number ordered. Back volumes are $7.00 each; back issues are $1.10 each except when more than five years old, and then $1.20 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Unsolicited manuscripts should be accompanied with return postage. Manuscripts, books and other materials for review, and correspondence regarding editorial and business matters should be addressed to the Publishers.

WARWICK AND YORK · *Publishers* · BALTIMORE, MD.

THE ACTIVITY MOVEMENT

By CLYDE HISSONG

In an attempt to overcome the weaknesses of the traditional school organization many progressive schools have developed new programs. These programs are so similar in character that collectively the changes have been referred to as the activity movement. This movement has claimed the center of the educational stage for a length of time sufficient to have engendered widespread interest in its outcomes and in its basic philosophy.

In Doctor Hissong's study an attempt has been made to discover the principles underlying the present activity movement, to determine the influence of traditional concepts in shaping the trends of the movement, and to see if in the light of the present knowledge of the child and his relation to his environment the movement rests upon a justifiable basis.

$2.00 plus 10¢ postage

WARWICK AND YORK

Publishers BALTIMORE

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXIX September, 1938 Number 6

LANGUAGE ACQUISITION BY DEAF CHILDREN AS RELATED TO HEARING LOSS AND AGE OF ONSET

NOEL KEYS

University of California

AND

LOUISE BOULWARE

San Francisco Public Schools

The major task confronting the deaf child seeking an education is unquestionably the acquisition of language itself. Yet the prediction of success at this presents uncommon difficulties. Performance tests of intelligence suitable for individuals with language handicaps appear of very doubtful value for prognosis. Thus Brown (1930) found the reading ability of pupils at the Jacksonville Illinois School for the Deaf to correlate $-.06 \pm .07$ with scores on the Arthur Point Scale of Performance Tests and $.003 \pm .07$ with the Pintner Non-Language Mental Test specially designed for use with the deaf.

Such results encourage a search for other bases of prediction, such as the age of onset of deafness and percentage of hearing loss or residual hearing. These data are available in the case of most pupils entering schools or classes for the deaf, but previous attempts to discover prognostic value in them have proved disappointing, as applied to children of entering school age.

PREVIOUS APPROACHES TO THE PROBLEM

In the census of 1910 (Deaf Mutes, 1917), children losing their hearing at any age up to eight were treated as having no advantage over those born deaf; and Reamer (1921, pp. 129–130) in standardizing an educational survey test on twenty-one hundred seventy-two deaf persons from eight to twenty-one years of age, found that, as compared with congenital deprivation, "onset of deafness before age six makes

no difference in later educational work; onset after six, however, makes a very significant difference and allows a child to progress much faster." Pintner (1933, p. 218) reached the more plausible conclusion that "those who retain their hearing up to age four or five do much better later on in tests involving language," and Fusfeld (1933, p. 201) reported for one hundred sixty-five students of Gallaudet College that of those born deaf or losing their hearing before two, only one in four exceeded the median in college work and on the Stanford Achievement Test, while one-half of those becoming deaf at ages two to four, and four-fifths of those losing their hearing after four attained this degree of success.

The findings with reference to degree of residual hearing are even more equivocal. Reamer (1921, p. 129) concluded "there is practically no difference between the mental ability or educational attainment of the partially and totally deaf, whether congenital or acquired." Her classifications were obtained without the benefit of our present audiometers, but two audiometric surveys of typical public-school populations have since revealed extremely low relationships with educational achievement. Waldman, Wade and Aretz (1930, pp. 151–152) discovered among five hundred fifty-two such pupils, after eliminating cases of irregular attendance, etc., that tested hearing showed a correlation of .23 $\pm$ .02 with Stanford EQ, and the same with IQ. Madden (1931, p. 50), however, found residual hearing by three audiometric tests to correlate but .16 $\pm$ .05 with reading, and .14 $\pm$ .05 with total Stanford Achievement scores; and the coefficients fell even lower after intelligence was partialled out. When fifty-nine definitely hard-of-hearing children were compared with a like number of normals, matched for Binet IQ and certain other factors, the differences fluctuated around zero.

In general, it would appear that the adventitiously deafened show a marked superiority in language learning over the congenitally deaf, though the age to which hearing must be retained in order to be of much permanent advantage is somewhat uncertain. On the other hand, the amount of residual hearing, as long as this does not exceed the limits of deafness, seems of no significance for the educator.

Even these conclusions, however, are open to question. The more carefully controlled of the studies from which they are drawn dealt only with hard-of-hearing children in regular school classes. It is doubtful whether findings so derived apply in any measure to the truly deaf, since the hard-of-hearing are accustomed from early childhood to think in verbal terms, while those deaf from infancy acquire

such symbols only by prolonged and painful effort. Moreover, the Stanford Achievement battery, like other tests for hearing children, presupposes a considerable familiarity with language for successful application. Whether a given deaf child will have attained this requisite proficiency depends in large measure on the nature and extent of the special instruction he has received. Inferences based on correlations with language status at a given age, regardless of the wide differences in educational history and background commonly found among the deaf, must be subject to grave reservations.

A much more convincing procedure would be to determine the progress made over a stated interval and under comparable educational opportunities, by individuals differing in amount of residual hearing and age of hearing loss. Such was the technique of the present investigation.

THE MEASUREMENT OF PROGRESS IN LANGUAGE

The study here reported was made possible by the development at the University of California of a series of tests of attainment in visual language. These were constructed specifically to measure the progress in language learning of deaf pupils during the first five or six years of instruction and prior to the level at which such batteries as the Stanford Achievement can be expected to function well. The standardization of these Visual Language Tests on the basis of the performance of three hundred fifty-nine pupils in the California School for the Deaf and special classes in the schools of three cities in that State is described in the manual by Keys and Pedersen (1938). The tests involved are of four types: word-knowledge, sentence-comprehension, paragraph-reading, and sentence-making. Each type has three parallel forms, save the sentence-comprehension, which has two; and each form yields a score in terms of age norms for deaf pupils.

When all the forms of each type are administered, reliabilities as computed by the method of split-halves and Spearman-Brown formula are extremely high. For pupils in a single grade, these range from .95 in the case of sentence-comprehension to .98 for word-knowledge, with .99 $\pm$ .001 shown for total scores on the entire series of eleven forms. The magnitude of these coefficients reflects in part the striking range of ability found in most classes for the deaf. Retests over the one-year interval of the present study, however, revealed correlations varying from .85 $\pm$ .02 for word-knowledge to .92 $\pm$ .01 for paragraph-reading, which must be considered highly satisfactory.

SUBJECTS AND PROCEDURE

In April and May of 1935, Keys and Pedersen, authors of the Visual Language Tests, administered the entire series to the California School for the Deaf at Berkeley, the Gough School of San Francisco, and the special classes for the deaf at Hawthorne School in Oakland. All classes were tested whose achievement was estimated to be at or below the level of hearing third-graders. In the Spring of 1936, within one week of the anniversary of its original testing, each class was retested by the writers of this article. Not all types of test were administered to each class, however, as the word-knowledge was known to be too simple for advanced pupils, and the paragraph-reading and sentence-making too difficult for beginners.

For purposes of the present study, a carefully defined selection was then made from the more than two hundred children tested. Only those pupils were included

Concerning whom the special school records indicated the age of onset of deafness as known, and who were there recorded as having been born deaf or become so before their sixth birthday.

Who had entered a special class or school for the deaf between the ages of four and one-half and seven, and continued under such instruction up to the time of testing.

Who were under fifteen years of age at the first testing (thus excluding a few mentally subnormal and incapable learners).

Who showed not more than sixty per cent of residual hearing in the better ear, this being regarded as the approximate dividing line between the deaf and the merely hard-of-hearing.*

In computing gains on any one of the four types of test, a further elimination was made of all children who had missed one or more of the forms of that type at either testing, or who had made a zero or perfect score on one, so that the full extent of improvement could not be accurately measured.

The number of cases available for the several computations after all eliminations ranged from forty-four and fifty-nine in the case of word-knowledge, to ninety-nine for the sentence-making tests. For gains in total Visual Language Age, only those individuals were considered who had taken every form of all four types of test in both

* Owing to the impossibility of determining accurately the percentage of hearing loss in many of the youngest children, fifteen cases of these whose audiometer tests proved unsatisfactory were nevertheless included in the correlation between age of onset of deafness and gains in word-knowledge.

years. Since no one type was given to all classes, only pupils near the middle of the grade range tested could qualify in this respect. Just twenty-five such were found, all of whom were residents of the California School for the Deaf. Of the other restricted groups, from seventy to eighty per cent were enrolled in that school.

Over ninety per cent of the pupils included were of ages seven to eleven inclusive in the course of the year covered by this study.

THE MEASUREMENT OF HEARING

Most of the tests of hearing loss were personally administered by Marshall Hester in the laboratory of the California School for the Deaf. First measurements were obtained with an audiometer of the 3*A*, or buzzer type. This yields a complex tone covering the frequencies typical of normal speech. This tone is varied in intensity until the threshold of hearing is ascertained. Percentages of hearing as thus determined are regarded as maximal. Children showing twenty-five per cent or more of hearing by this criterion were given a complete radio-ear test with an instrument similar to the usual 2*A* audiometer. Residual hearing at five frequencies distributed over the range of speech sounds was then averaged.

Both audiometers named are of the individual type, and yield a separate score for each ear. All pupils included in the calculations of the present study showed grave impairment in both ears, and the difference between the two was seldom great. The record of the better ear was taken as the index of residual hearing, since Madden (1931, pp. 50–51) found this to yield correlations uniformly superior to those for the poorer ear or average of two ears, and within .01 of the multiple *R* obtainable from both ears.

To determine the retest reliability of the audiometer scores for the group studied was entirely impractical, but this was estimated to be of the order of .8 for the 3*A* records which were the determining factor in the dichotomy employed (cf. Upshall, 1929, p. 56).

SIGNIFICANCE OF THE AGE AT WHICH DEAFNESS OCCURRED

Calculation of agreement in terms of the product-moment coefficient of correlation, *r*, assumes normality of distribution in each of the functions correlated. Since fully half of the pupils in the selected sample were recorded as deaf from birth, this assumption could not be justified with reference to the age at which deafness occurred. The reliability of differences between groups differing in age of onset was, therefore, employed instead. However, in order to indicate the degree

of relationship for those accustomed to thinking in terms of correlation coefficients, the table showing critical ratios has been followed by one giving the corresponding bi-serial correlations (Garrett, 1937, pp. 366–371). Bi-serial r's so found are presumably comparable in magnitude to the product-moment r's, which would have been obtained had the data lent themselves to such treatment.

As stated, no pupil was included in the restricted groups who was six years or older at the onset of deafness. The importance of hearing up to that age has been amply demonstrated by Reamer and confirmed by Fusfeld, as cited. The question raised in the present investigation

TABLE I.—NUMBER OF CASES EMPLOYED IN COMPUTING RELATIONS BETWEEN PROGRESS IN LANGUAGE LEARNING AND AGE OF ONSET OF DEAFNESS

Type of test	Years of age at onset		Total
	Under two	From two to five	
Word-knowledge	41	18	59
Sentence-comprehension	47	20	67
Paragraph-reading	53	21	74
Sentence-making	74	25	99
Visual Language Age	16	9	25

TABLE II.—DIFFERENCES IN MONTHS OF PROGRSS SHOWN ON VISUAL LANGUAGE TESTS OVER ONE YEAR OF INSTRUCTION BY GROUP 1 (HAVING LOST THEIR HEARING AT TWO TO FIVE YEARS OF AGE) AND GROUP 2 (DEAF BEFORE TWO)

Type of test	Average gain of group		Difference $\text{mean}_1 - \text{mean}_2$	$\sigma_{diff.}$	Critical ratio, or Diff./$\sigma_{diff.}$
	1	2			
Word-knowledge	12.33	12.68	−0.35	2.46	−0.14
Sentence-comprehension	11.70	11.04	0.66	1.71	0.39
Paragraph-reading	7.14	9.21	−2.07	1.99	−1.04
Sentence-making	11.96	9.19	2.77	2.65	1.05
Visual Language Age	10.28	12.53	−2.25	3.12	−0.72

concerns the extent to which hearing retained to the age of two to five years is likely to give a lasting advantage in later language learning over children whose deafness dates from birth or before their second birthday. The dividing line was here set at two, since the normal child of two or over will have attained at least the beginnings of language.

TABLE III.—BI-SERIAL CORRELATION BETWEEN GAINS IN VISUAL LANGUAGE SHOWN OVER ONE YEAR OF SCHOOLING AND AGE AT ONSET OF DEAFNESS

TYPE OF TEST	BI-SERIAL r PE
Word-knowledge	$-.02 \pm .04$
Sentence-comprehension	$.04 \pm .11$
Paragraph-reading	$-.13 \pm .03$
Sentence-making	$.15 \pm .09$
Visual Language Age	$-.16 \pm .17$

The number of cases in each group employed in the several computations is shown in Table I; the differences and their critical ratios, in Table II; and the corresponding bi-serial r's, in Table III. A difference is considered "practically certain," or too great to be explained by chances of sampling, when it amounts to three or more times its standard error. A glance at the critical ratios in the final column of Table II will show that in no instance do they approach this figure. The largest differences barely exceed one sigma, which would occur almost one time in three by chance alone. Nor does either group consistently excel the other; the algebraic average of the ratios is almost zero. The bi-serial r's point to the same conclusion.

So far from possession of hearing until two to five years of age leading to more rapid progress in later schooling, the differences noted are strictly of the order of chance, with the group born deaf or deafened before two slightly excelling on three comparisons out of five. This would appear to support the observations of Reamer as against those of Fusfeld, who found students having lost their hearing at two to four appreciably surpassing those handicapped from infancy. Direct comparison is difficult, however, since the present study is based on improvement shown over a single year, rather than status attained.

As a check upon the latter, Visual Language scores and ages at the time of second testing were noted for the twenty-five members of the selected group having complete records on all eleven forms of the tests. Of these, the sixteen known to have been deaf by their second birthday averaged ten years, seven months of age and scored a mean Visual Language Age of nine years, seven months, showing an average quotient of 91 ± 2.3. The nine who had lost their hearing between two and five, however, averaged only nine years, four months of age on the date in question, and scored nine years, four months on Visual Language, showing an average quotient of 100 ± 3.3. The difference of nine points in Visual Language Quotients amounts to only 2.3 times its probable error, and so is far from reliable. In all likelihood, however, it points to a real superiority on the part of the adventitiously deafened. We may then infer that, while children losing their hearing

at two to five years of age will not usually be making more rapid *progress* in language around the ages of eight to ten than those deaf from infancy, they may nevertheless be expected to show a slightly more advanced *level* of attainment.

INFLUENCE OF RESIDUAL HEARING ON LANGUAGE LEARNING

In investigating the relation between residual hearing and improvement in language, the line between the totally and partially deaf was set at twenty per cent of normal hearing, or eighty per cent loss in the better ear. Thus Knudsen (1934, p. 48) states that "those who have less than twenty per cent of normal hearing may be regarded as totally deaf, having lost all serviceable hearing. There are not many who are rated as totally deaf who have less than ten or fifteen per cent of hearing." The partially deaf were defined as those having from twenty to sixty per cent of normal hearing in the better ear. These comprise pupils who can profit to some extent from the use of hearing aids, but cannot be restored to a normal hearing life by their assistance.

TABLE IV.—NUMBER OF CASES EMPLOYED IN COMPUTING THE RELATION BETWEEN PROGRESS IN VISUAL LANGUAGE AND PERCENTAGE OF RESIDUAL HEARING

Type of test	Number		Total
	Totally deaf (0 to 20 per cent hearing)	Partially deaf (21 to 60 per cent hearing)	
Word-knowledge	25	19	44
Sentence-comprehension	40	27	67
Paragraph-reading	40	29	69
Sentence-making	58	41	99
Visual Language Age	14	11	25

The number of cases employed in the computations for the several types of test is shown in Table IV. It will be seen that the majority of the groups selected were totally deaf in the sense indicated.

Differences in the progress made by the partially as compared with totally deaf are indicated in Table V. These findings are in striking contrast to those of Table II. Whereas the average gains recorded by the totally deaf range from 7.9 to 10.1 months, those of the partially deaf vary from 10.8 to 17.3 months, with no overlapping in the means. Moreover, the critical ratios are sufficient to afford "practical certainty" of the reality of these differences in four out of the five comparisons.

TABLE V.—DIFFERENCE IN MONTHS OF PROGRESS SHOWN BY TOTALLY AS COMPARED WITH PARTIALLY DEAF PUPILS ON VISUAL LANGUAGE TESTS OVER ONE YEAR OF INSTRUCTION

Type of test	Average gain by		Difference $mean_1 - mean_2$	σ_{diff}	Critical ratio, or Diff./$\sigma_{diff.}$
	1 Partially deaf	2 Totally deaf			
Word-knowledge.....	14.90	10.08	4.82	1.15	4.19
Sentence-comprehension....	15.37	8.45	6.92	1.48	4.68
Paragraph-reading.....	11.45	6.80	4.65	0.55	8.45
Sentence-making.....	10.78	9.26	1.52	1.13	1.35
Visual Language Age.....	17.27	7.87	9.40	2.50	3.76

Bi-serial correlations for the same comparisons are presented in Table VI. The first column indicates the bi-serial coefficients as computed, together with their probable errors. It will be noted that those for sentence-comprehension and total Visual Language Age are highly significant. In the final column the correlations are shown after correction for attenuation by use of the formula,

$$r_{corrected} = \frac{r_{gains,\ hearing}}{\sqrt{r_{gains} \cdot r_{hearing}}},$$

taking the reliability of the audiometer records as .80. The writers are indebted to Professor H. S. Conrad of the University of California for derivation of the following formula for the reliability of gains between tests 1 and 2,

$$r_{gains} = \frac{r_{11} - r_{12}}{1 - r_{12}}.$$

Here r_{11} represents the reliability coefficient of the particular type of test, and r_{12} the correlation between scores on this test in 1935 and those in 1936.

From Tables V and VI it is clear that differences in residual hearing among the deaf show a substantial relation with their improvement in language over a year period, according to four of the five criteria employed. That this does not hold true for scores on sentence-making as well may be owing to the fact that the tasks of that type proved too difficult for many of the pupils tested.

As regards Visual Language Age on the eleven forms combined, the correlations, either with or without correction for attenuation, compare with those commonly found in the prognosis of achievement

TABLE VI.—BI-SERIAL CORRELATION BETWEEN GAINS IN VISUAL LANGUAGE OVER ONE YEAR OF INSTRUCTION AND AMOUNT OF RESIDUAL HEARING (PARTIAL *vs.* TOTAL DEAFNESS)

Type of test	Bi-serial *r* as obtained	Reliability of test employed	Correlation between 1935 and 1936 scores	Reliability of gains	Coefficient of correlation corrected for attenuation
Word-knowledge.....	.30 ± .15	.99	.85[1]	.92	.35
Sentence-comprehension..	.40 ± .09	.97	.90	.65	.55
Paragraph-reading.....	.30 ± .10	.98	.92	.67	.42
Sentence-making.....	.09 ± .09	.96	.89	.69	.12
Visual Language Age.....	.63 ± .12	.99	.85[1]	.95	.72

[1] The lower correlations over a one-year interval in the two cases noted are attributable chiefly to the greater homogeneity of the pupils in the restricted groups concerned.

in school subjects on the part of hearing pupils. Certainly the prediction afforded is of a very different order from those discovered in previous attempts to forecast language learning by deaf children. Moreover, analysis of the performance of the twenty-five boys and girls in this group reveals that the superiority of the partially hearing was by no means confined to progress over the year studied, but characteristic of their total achievement to date. The eleven partially deaf pupils showed in 1936 an average Visual Language Age of ten years, two months and a like chronological age, making an average quotient of 100 ± 3.4. The fourteen totally deaf, on the contrary, recorded a mean Visual Language Age of nine years, no months, with a chronological age of ten years, one month, making an average quotient of 89 ± 2.4. Owing to the small number of cases, the difference amounts to only 2.5 PE. It is worthy of note, however, that this difference was obtained despite the fact that no one of the eleven partially deaf involved had more than forty-eight per cent residual hearing in the better ear, according to the 2*A* audiometer, while four of the fourteen "totally deaf" recorded twenty per cent hearing on the 3*A* type of instrument.

SUMMARY AND CONCLUSIONS

The present study treats of factors in the acquisition of language by children, all of whom may properly be classified as deaf rather than

merely hard-of-hearing. Not only have the non-verbal tests of intelligence applicable to such children proved of little service in predicting their success in language learning, but such previous studies as are available would indicate that differences in degree of deafness and age of onset are also of slight prognostic significance when the disability dates from younger than five.

As regards age of onset, only one American investigator has reported persons losing their hearing at from two to five years as showing any appreciable lasting advantage from having once heard, and his observation was made on college students. The present study found pupils of this description in no way surpassing those deaf by the age of two in improvement shown over one year of instruction on the elementary-school level. Analysis of a restricted group of twenty-five cases suggests that children deafened at two to five will excel those deaf from infancy in level of attainment, even though not in rate of progress; but the observed difference is statistically insignificant, the number too small, and the uncontrolled variables too numerous to warrant a conclusion on this point. In general, our results tend to support the early observation of Pintner and Paterson (1917) as to the comparatively slight permanent advantage from hearing which is lost much below the age of entrance to school.

The findings with reference to percentage of residual hearing are in sharp contrast to the above. Groups of partially deaf pupils with a range of twenty-one to sixty per cent and a median of well under forty per cent residual hearing in the better ear surpassed on every point of comparison the totally deaf having not to exceed twenty per cent hearing in either ear. As regards progress over one year of instruction, differences between these groups were significant by four of the five test criteria; and a sampling of twenty-five children showed a superiority of more than ten points in average Visual Language Quotient of the partially hearing.

Bi-serial correlations between residual hearing and gains in total Visual Language Age amounted to .63 before, and .72 after correction for attenuation. While too low for accurate individual prognosis, these coefficients indicate that extent of hearing loss affords a more valid basis for forecasting language acquisition on the part of deaf children than any criterion which the writers have discovered in the existing literature. To judge from the present findings, school authorities would be justified in utilizing the percentage of residual hearing, as determined by careful audiometric examination, to classify deaf pupils entering upon language instruction in the same manner and

with nearly the same degree of confidence that intelligence quotients or similar measures of scholastic aptitude are employed in sectioning hearing first-graders. The audiometer tests which are essential for medical diagnosis can thus serve for educational prognosis as well.

The question arises as to how this conclusion can be reconciled with the failure of the one earlier study (Reamer, 1921) to discover a difference between the language attainments of partially as opposed to totally deaf persons. In the judgment of the writers, this discrepancy in findings is attributable primarily to the great superiority of the instruments employed in the present investigation for the measurement both of language achievement and of hearing loss. Important contributory factors may be found in the heavier emphasis now placed on oral and written language as contrasted with manual signs in the education of the deaf, and the better utilization of comparatively small amounts of residual hearing made possible by such instructional aids as the "radio ear."

REFERENCES

———: "Deaf Mutes in the United States. Analysis of the Census of 1910." *American Annals of the Deaf*, Vol. LXII, January, 1917, pp. 62–82.

Brown, A. W.: "The Correlation of Non-Language Tests with Each Other, with School Achievement, and with Teachers' Judgments of the Intelligence of Children in a School for the Deaf." *Journal of Applied Psychology*, Vol. XIV, August, 1930, pp. 371–375.

Garrett, H. E.: *Statistics in Psychology and Education.* Revised Edition. New York: Longmans, Green and Co., 1937.

Keys, Noel, and Pedersen, Virginia D.: *Visual Language Tests* and *Manual.* Chicago: Plymouth Press. To appear shortly.

Knudsen, V. O.: "Hearing Aids." In *Fifth Yearbook* of the California Council for the Education of Exceptional Children. Berkeley, California: State School for the Deaf, 1934, pp. 47–56.

Madden, R.: *The School Status of the Hard of Hearing Child.* Teachers College, Columbia University, Contributions to Education, No. 499, 1931, pp. 64.

Pintner, Rudolf: "Contributions of Psychological Testing to the Problems of the Deaf." *Proceedings* of the International Congress on the Education of the Deaf. West Trenton, N. J.: School for the Deaf, 1933, pp. 213–220.

Pintner, R., and Paterson, D. G.: "Psychological Tests of Deaf Children." *Volta Review*, Vol. x, December, 1917, pp. 661–667.

Reamer, Jeanette C.: "Mental and Educational Measurements of the Deaf." *Psychological Monographs*, Vol. XXIX, No. 3, 1921, pp. 130.

Upshall, C. C.: *Day Schools versus Institutions for the Deaf.* Teachers College, Columbia University, Contributions to Education, No. 389, 1929, pp. 104.

Waldman, John L., Wade, F. A., and Aretz, C. W.: *Hearing and the School Child.* Philadelphia: Temple University, 1930, pp. 223.

EFFECT OF AGE ON THE SPEED-POWER RELATIONSHIP WITH REFERENCE TO TESTS OF INTELLIGENCE

RALPH R. BROWN

Psychologist, United States Public Health Service

In connection with the construction of a mental alertness test for application to drug addict and penitentiary populations, a question arose concerning the value of speed tests as compared with power tests. In such populations, which have an average age close to thirty-six, with an approximate age range from sixteen to seventy-five, it seemed justifiable to question the use of time-limit tests which have been standardized on younger groups.

The evidence in the literature bearing upon this question is somewhat conflicting, but the indications are that with advancing age there is a decrease in the speed function, as well as a relative change in ability with respect to certain types of test material. The authors of the original Army Alpha examinations[5] report that a doubling of the time limits of the Alpha subtests did not materially affect the original rankings. Jones and Conrad[3] found that elimination of the speed factor in the Alpha did not affect the developmental curve. Other workers, however, report significant differences between old and young adults with respect to tests in which there are speed components. Lorge[4] found that on the Army Alpha, Bregmann Revision, Wells Revision, Otis Self-administering, and the Thorndike Intelligence Examination for High-School Grades, older adults (age forty and over) scored significantly lower than a young group (age twenty to twenty-five), although both groups were equal with respect to the CAVD tests, which are power tests, with no time limit. Both W. R. Miles[7] and C. C. Miles[6] report a decline in ability with age, which they believe is due in large part to the influence of a speed factor.

Jones and Conrad[3] have reported that subjects in the sixth decade of life derive forty per cent of their total Army Alpha score from subtests 4 and 8 (Word Opposite and General Information). Foster and Taylor[2] have shown that decline in ability with age is to a large extent a function of the type of test material. These authors found an increase in score with advancing age in such functions as comprehension of questions, detection of absurdities, and definition of abstract terms; whereas there was a decrease in tests of drawing from memory, dissected sentences, placing three words in a sentence, and number of words given in three minutes.

The purpose of the present study was to determine the effect of age upon the speed-power relationship with respect to a non-language

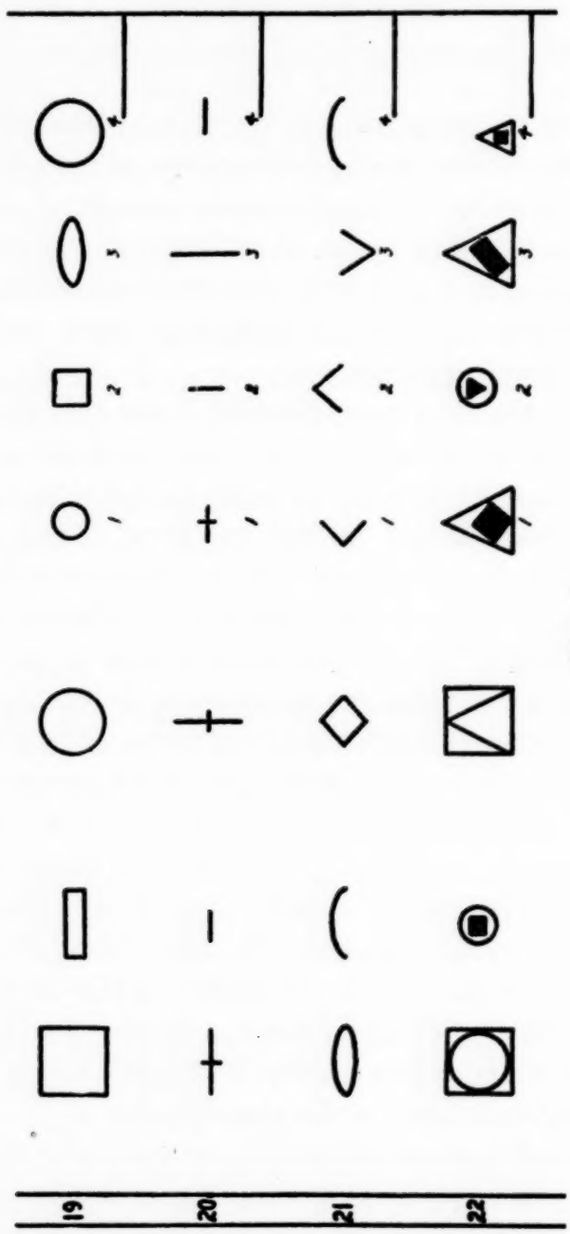

FIG. 1.

analogies test, which is being considered as a component of a mental alertness test now under construction.

PROCEDURE

A symbol-analogies test was constructed containing one hundred sixteen items ranging from easy to difficult (subjectively judged), a few items from which are reproduced in Fig. 1.

The test was multilithed and administered without time limits to a group of subjects (drug addicts who had been off drugs for more than four months), who were distributed symetrically (ten-twenty-forty-twenty-ten per cent) with respect to intelligence as measured by the original Army Alpha examination. From these results a thirty-one-item scale was constructed, the items of which increased by equal increments from easy to difficult. The short scale correlated .95 with the original test, with a split-half reliability of .91. This test will be referred to as the power test.

Another test was constructed by selecting fifteen items of approximately equal difficulty. The test was doubled in length by duplicating each item, making a test of thirty items, with the arrangement such that duplication of items did not begin until after the first fifteen items were completed. Reliability for this test was .94 by the split-half method with a time limit of one and one-half minutes. This test will be referred to as the speed test.

These two tests were administered to a new group of one hundred thirty addicts who were off drugs at least four months, selection being made on the basis of age and normal intelligence distribution. Forty-four of the subjects were below thirty, forty-one were above fifty years of age, and forty-five were in the middle-age group (thirty-one to fifty). The power test was given first, without time limit, followed by the speed test.

RESULTS

The general results are presented in Table I, with the critical ratios shown in Table II. It may be noted from these tables that the groups are fairly equal with respect to the Army Alpha total score and the number correct on the power test. On the speed test, however, the young group has a mean of 9.90 (number correct) as compared with 7.80 for the old group. The critical ratio here is 1.39. It may also be observed that the standard deviation of the speed test scores for the young group is larger than that of their power test, whereas the reverse condition holds for the old group; *i.e.*, the standard deviation of the speed test is less than their power test.

Apparently when applied to older adults, the speed test tends to restrict the scatter; whereas on younger subjects, this effect is not noticed.

Expressed in terms of overlapping scores, fifty-seven per cent of the young group equal or exceed the mean of the old group on the speed

test; whereas thirty-nine per cent of the old group equal or exceed the mean of the young.

TABLE I.—MEANS AND STANDARD DEVIATIONS ON POWER, SPEED, AND ARMY ALPHA TESTS, BY AGE GROUPS

	Young *N* 44		Middle *N* 45		Old *N* 41		Total group *N* 130	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	25.61	3.00	39.13	5.04	56.46	5.05	40.02	13.24
Power test items correct	12.84	6.23	12.00	6.72	11.93	6.80	12.28	6.59
Speed test items correct	9.89	7.83	8.96	6.23	7.80	6.10	8.91	6.82
Total Alpha (less subtest I)	57.34	33.33	58.91	34.77	55.93	26.47	57.44	31.87

TABLE II.—STATISTICAL SIGNIFICANCE OF MEAN DIFFERENCES EXPRESSED IN TERMS OF CRITICAL RATIOS[1]

	Power test	Speed test	Total Alpha score
Young—middle	.60	.63	−.22[1]
Young—old	.64	1.39	.22
Middle—old	.05	.87	.45

[1] Minus sign indicates that the mean of the second-named group is larger.

The following correlations are of significance in evaluating the effect of age on the speed-power relationship with reference to the symbol-analogies test:

Speed-power correlation—young group	.79
Speed-power correlation—middle group	.71
Speed-power correlation—old group	.65
Speed test with age (total group)	−.16
Power test with age (total group)	−.07

It is apparent that under the present conditions there is an indicative, but not statistically significant, difference between the old and young groups with reference to the speed factor. The difference is, however, of sufficient significance to warrant consideration when applied as part of a test battery in which other subtests may show the same trend, thus resulting in an accumulative effect. In the Army Alpha, for example, subtests 2, 3, 5, 6, and 7 all give low negative cor-

relations with age, a condition which would significantly affect the total score if it were not for the positive correlations with age on subtests 4 and 8. Even with this balancing effect of subtests 4 and 8, however, the Army workers (5, p. 817) found a correlation of negative .192 between age and total Alpha score on a group of five thousand four hundred four medical officers. Lorge[4] reports a negative .408 correlation between age and the total Army Alpha score, and refers to this test as one which measures power and speed in undifferentiated amounts.

The present results, when interpreted in light of the findings reported in the literature, indicate an answer to the question stated at the beginning of this paper. It is apparent that power tests are more suitable for application to older persons, although a measure of the speed factor would prove of diagnostic value in many instances. Those tests which measure speed and power in undifferentiated amounts are to be avoided.

In a survey of psychometric tests being administered in Federal and state penal institutions,[1] the writer found that the most frequently used test was the Army Alpha, or one of its revisions. Next came the Army Beta and the Otis, and three institutions were using only individual tests, generally the Binet. None of these tests allow for isolation of the speed from the power factors, although the Binet offers some opportunity to note gross differences at certain levels.

Although from the standpoint of administration it is obvious that a test with unlimited time would prove impracticable, nevertheless it should be possible to select time limits for subtests of a battery which would handicap only a very few individuals, and these only to a limited extent. A time limit of twenty minutes on the power test described in this article, for example, would prove only a slight handicap to a few individuals. The gradations of a scale which attempts to cover the range of adult intelligence must increase rapidly from easy to difficult, with the consequence that the majority of subjects rapidly reach the limit of their capacity on each subtest, assuming that the test items are not too finely graded with respect to increments of difficulty. Such a scale, of course, would be unsuited for determinations of fine differences at any particular intellectual level, but this is generally not essential insofar as the function of a test for penitentiary populations is concerned.

The speed factor in intelligence is, however, of significance with respect to occupational adjustment, and consequently it would seem

advisable to include a few measures of mental speed in any intelligence test battery which is to be applied to adults. Further, it is quite possible that a knowledge of the relationship between speed and power may prove of benefit in personality classification beyond the purely intellectual factors.

SUMMARY AND CONCLUSION

Speed and power tests for non-verbal analogies were constructed and applied to a drug addict population containing old, middle, and young age groups. The results indicated a tendency for the young group to exceed the old on the speed test, although both groups were about equal on the power test. When interpreted in light of findings reported in the literature, the results offer further substantiation of a true difference between old and young persons with respect to the speed-power relationship.

The need for differentiation of speed from power factors has been discussed with reference to application of intelligence tests to penitentiary and drug addict populations.

LIST OF REFERENCES

1. Brown, Ralph R.: *A Survey of Psychometrics in State and Federal Penitentiaries for Adult Males.* Unpublished Report, United States Public Health Service, August 1934.
2. Foster, J. C. and Taylor, G. A.: "The Applicability of Mental Tests to Persons over Fifty Years of Age." *J. Appl. Psy.*, Vol. IV, 1920, pp. 19–58.
3. Jones, H. E. and Conrad, H. S.: "The Growth and Decline of Intelligence. A Study of a Homogeneous group between Ages of Ten and Sixty." *Genet. Psy. Monog.*, Vol. XIII, 1933, pp. 223–298.
4. Lorge, Irving: "The Influence of the Test upon the Nature of Mental Decline as a Function of Age." *Jrn. Ed. Psy.*, Vol. XXVII, No. 2, Feb., 1936, pp. 100–110.
5. *Memoirs of the National Academy of Sciences*, Vol. XV, 1921. Government Printing Office.
6. Miles, C. C.: "Influence of Speed and Age in Intelligence Scores of Adults." *J. Gen. Psy.*, Vol. X, 1934, pp. 208–210.
7. Miles, W. R.: "Abilities of Older Men." *Person. J.*, Vol. XI, 1933, pp. 352–357.

MEANING AS A FACTOR IN LEARNING*

H. B. REED

Fort Hays, Kansas, State College

The psychology of learning has been dominated by the concept of associationism, with its explanatory principle of contiguity, and by the concepts of animal learning with its explanatory principle of the conditioned reflex. Ebbinghaus,[11] who is the father of experimental investigations of learning as well of the higher thought processes, accepted the concepts of associationism and worked out his experiments on their foundation. For him learning was essentially an activity of memorizing descrete units of verbal material. The process of imprinting was repetition; and the combining principle, contiguity. Under his influence an immense impetus was given to the investigation of numerous quantitative factors influencing the learning and retaining of meaningless material, such as the number of readings, the number of units, the distribution of the readings in time, the distribution of the readings in space, order and position of the units, rhythmic presentation, rate of presentation, sensory mode, retroactive inhibition, proactive inhibition, length of interval between learning and relearning, organic state of the learner, attitude and motivation of the learner, physical and other environmental conditions, and age, sex and ability of the learner.

Before these investigations had gone very far, similar problems began to be studied in motor learning in humans and in animals. The pioneer investigations of Bryan and Harter[8] and of Thorndike[33] gave a special impetus to the study of learning curves and of the numerous factors that influence their shape. In these studies the maze became the most important psychological instrument. It is the motor parallel of a row of nonsense syllables. It is meaningless, quantitatively variable, and a collection of descrete units. The learning of it requires the association of these units into a certain order, it is mastered primarily by repetition, and the combining principle is contiguity. The maze, first used in animal experiments, is now used with almost equal frequency in problems of human learning. Like a row of nonsense syllables, it makes much of the learning process objectively observable, makes the control of variables and of experimental condi-

* An address delivered before the Psychological Division of the A.A.A.S., Denver, Colo., June 22, 1937.

tions comparatively easy, and yields quantitative results. For these reasons it has been widely used.

The list of experimental studies in learning made with mazes, nonsense syllables, and other meaningless materials constitute a bibliography of many pages. In method they are, for the most part, scientific, and the results are objective and quantitative. The results are valuable so far as they go, but they are open to certain objections. They have over-simplified the learning process, covered a too narrow field, favored mechanistic interpretations, and kept the principal field of human learning, that of meaningful materials, largely untouched. It is true that meaningful materials were occasionally studied, but, when this was the case, they were used principally for purposes of comparison to show the influence of another variable; namely, the quality of the material. But the studies which sought to investigate logical learning on its own account are very few. Because of this fact the practical contributions of the psychology of learning are rather disappointing. It is the belief of the writer that this condition will be radically changed when once we begin to investigate seriously logical learning for its own sake.

ADVANTAGES OF RECOGNIZING MEANING

The first notice of meaning as a factor in learning observed by the writer[26,27] was in an article on "Paired Associates," which was published in 1918. In this experiment, the writer observed that certain pairs of words were easily learned, while others were learned and remembered with great difficulty. Upon investigation it was found that pairs which were quickly learned had some meaning for the learner or were associated in some way or other by him. Those that were learned with great difficulty had no meaning for the learner and could not be associated by him. When the retention of such pairs was tested after one day, it was found that they were almost invariably forgotten. From the results the writer concluded that materials which may be associated or which have meaning are quickly learned and well remembered, while those which cannot be associated and which are, therefore, meaningless are slowly learned and quickly forgotten. In a later experiment,[28] published in 1924, some definite advantages of the perception of meaning in learning were discovered. In this experiment the writer demonstrated that the perception of meaning greatly increased the quantity of material which can be grasped in one act of thought, that it greatly increased the speed and ease of learning, and that it

greatly increased the amount retained. The influence of meaning upon quantity of material that can be grasped in one act of attention was shown in memory-span experiments. The memory span for college students for disconnected words was found to be 6.55; while for words in meaningful sentences it was found to be 25.36. Here we find that the memory span for words in meaningful sentences to be four times as great as for words in disconnected form. The influence of meaning on ease of learning was shown by the difference in the time required to learn two selections of prose equal in length but different in difficulty. The meaningful prose required 111.25 seconds to memorize, while the difficult prose required 261.25 seconds. Similar results were found in an experiment made by Lyon,[22] who required ninety-three minutes time to learn two hundred nonsense syllables, but only ten minutes to learn two hundred words of poetry. The influence of meaning on retention was shown in the number of ideas retained after two weeks, from the selections entitled "Origin of Ideas" and "Marble Statute." Of the former, which consisted of difficult and abstract prose, only four ideas were retained; while of the latter, which was easy meaningful prose, 39.3 ideas were retained, or almost ten times as many. According to these experiments, if material is meaningful to the learner, it is easily learned and long remembered; while if it is meaningless, it is difficult to learn and difficult to remember. If this is true, it has some evident application to teaching, and also means that drill procedures, which have hitherto been emphasized as important economical learning, have relatively less importance.

The perception of meaning is an important aid in the transfer of training. Ruger[30] in his experiments with puzzles found that transfer from one puzzle to another involving the same principle was greatly facilitated if the learner discovered the principle involved in the first puzzle. Judd and Scholkow[19] found that skill in shooting at a target in one depth of water was more readily transferred to shooting at a target in another depth in case of boys who were taught the principle of the refraction of light. J. C. Peterson,[23] who experimented with the ability to transfer training from one drawing game to another, which involved the same formula, found that subjects who discovered the correct formula in the first game saved eighty-nine per cent of the trials in solving the second game, while subjects who failed to discover the correct formula saved only thirty-one per cent of their trials in the same situation. All these facts indicate that the perception of a principle involved in two or more activities greatly facilitates the transfer of

skill from one to the other. The main conclusion from these studies on the advantages of the perception of meaning, is that it contains the secret of economical learning. This means that in order to learn easily, retain well, and transfer skill effectively, we should emphasize the importance of the perception of meanings, relationships, and principles.

RELATION OF MEMORY OF MEANINGFUL MATERIAL TO OTHER FACTORS

Investigations of meaningful prose are spoken of in psychological literature as experiments on prose substance, while those on meaningless material are called rote memory. Memory for prose substance is thought by some psychologists to be totally different from memory for rote learning, but it probably differs from it in degree rather than in kind. That memory for prose substance is different psychologically from rote memory is indicated by the enormous differences in the time and effort required to learn the two kinds of material, by the low correlations between these two functions, and by the differences in the shape in the curves of retention for the two kinds of material. In the foregoing paragraphs statements have already been made showing that meaningful material may be memorized in one-ninth of the time which is required to learn the same quantity of meaningless material. These immense variations point to the wide difference in the functions involved. A number of investigators have calculated correlations between rote memory and memory for prose substance. Bennett,[6] for example, found the correlation to be only −.01, while Anastasi[1] found it to vary from .094 to .154. Other investigators, such as Jones and English,[18] and English, Welborn, and Killian,[12,13] found the correlation to vary from .328 to .419. Even the highest of these correlations does not show much relationship between the two types of memory, while the lowest shows none at all. These variations are possible because of the amount of difference in meaning between the connected and disconnected materials used. Some connected material may be meaningful and easy, while in another the meaning may be difficult to perceive. On the other hand, some disconnected material suggests no associations while another suggests a large number of them. On the basis of correlation studies, however, we can at least see that it is very dangerous to draw any applications from the results obtained from learning non-meaningful material to the learning of meaningful material—a habit which is all too common among psychologists.

The results on the curve of retention for the two types of memory are not consistent. The typical curve of forgetting is one discovered by Ebbinghaus,[11] which shows extremely rapid forgetting for the first day and then slow forgetting thereafter. According to Ebbinghaus, who measured retention by the saving method, forty-two per cent of material that was completely learned is forgotten after twenty minutes; after one day, sixty-six per cent is forgotten; and after thirty days, eighty per cent is forgotten. The curve for this material shows an almost vertical drop for the first day, after which it turns to the horizontal and declines slowly. Probably the most careful investigation of forgetting meaningful material was made by Dietze,[10] who had students read three selections of meaningful material, and then measured retention by the per cent of ideas retained, by multiple choice tests. The shape of the curve of forgetting depended upon the method of measurement. According to one method of measurement, the curve was similar to the Ebbinghaus curve, but showed greater amounts retained for the intervals investigated; according to another method, the curve showed a gradual decline from ninety per cent retention immediately after learning to fifty-two per cent thirty days after. Exact comparisons are difficult to make because of the differences in the method of measurement. If we compare the figures regardless of the method by which they were obtained, Dietze showed as much retention after thirty days, of material which had been read once, as Ebbinghaus after one day, of material which had been completely memorized. In another sense, however, no comparison is possible, because Ebbinghaus could consciously recall nothing, even after one day. Dietze's subjects consciously recalled a little over half of the material learned, after an interval of thirty days. If retention for the two types of memory is measured in the same way, there is considerable resemblance in the two curves. This leads the writer to believe that the two types of memory are different in degree rather than in kind, but, whatever method of comparison is used, there is little doubt that the amount retained of meaningful prose is far greater than for meaningless material.

Another investigation of the retention of prose substance is the one made by English, Welborn, and Killian.[12,13] They measured two types of retention of the same material, verbatim items and ideas. The former required the recognition of verbatim phrases and factual tests, and the latter, the recognition of meanings expressed in paraphrases and summaries. The most remarkable result of this investigation was the difference in the curves for the two types. The verbatim

items showed the usual decline in relation to the length of the interval between learning and testing. The paraphrases and summaries, on the contrary, showed an increase in relation to time up to intervals of seventy-one days. This divergence in the two classes of curves is another proof that the memory for prose substance is radically different from rote memory.

The relation of memory for prose substance to mental age and sex was investigated by Dietze. He found a gradual increase in the amount remembered from one reading, from Grade VII to Grade XII. These results indicate increase in ability to remember ideas, with age during the period of growth. When, however, the factor of chronological age was isolated by means of partial correlations, it was found to have no relation to memory for prose substance. Mental age, however, had a correlation of .55 with prose substance, a figure that was very little reduced by taking out the influence of the other factors by the partial-correlation technique. This is what we should expect, for the amount of meaning that is perceived from given material is to some extent a function of the maturity of the learner's mind. With respect to sex differences in memory for prose substance, Dietze found the boys to be superior to girls; but Shaw[31] and Pyle[25] found a superiority in favor of the girls. Probably, these sex differences are due to the quality of the material rather than to any native sex characteristics of the learners.

The relation of memory of prose substance to the number of repetitions is very different from that for rote memory. Ebbinghaus, who was the first to investigate this relationship, found that for every three repetitions in learning, he saved one repetition in relearning. Other investigators have found that the relationship is less direct, but all agree that retention in improved by increasing the number of repetitions, and that the upper limit, beyond which further repetitions are unprofitable, is very high. Similar investigations for prose substance are very limited.

In the study by English, Welborn, and Killian, the effect of from one to five repetitions on verbatim memory and on prose-substance memory was investigated. It was found that the verbatim items showed consistent gains up to two repetitions, but no consistent results were found for the retention of prose substance. Similar results were found by Hunsley[17] in the writer's laboratory. She investigated the relationship of the number of readings to the amount retained of prose selections having over a thousand words. There were increments in

the amount retained for such additional repetitions up to five, but the size of the increments beyond three readings was negligible. This indicates that memory for prose substance is very little influenced by number of readings. From this it is clear that memory for prose substance does not depend so much on the fact of repetition as on other factors, the most important of which is probably comprehension.

What is true of the number of repetitions, appears also to be true of their distribution and form. The experiments of Foster,[14] Austin,[2] and Welborn[12] indicate that certain spacings are better than others, but that the differences are much less marked than in meaningless material and in motor learning. The experiment of Gates[15] is evidence that the recitation method of memorizing is far more effective for the retention of meaningless than of meaningful material. Such results again suggest that it is unsafe to formulate the law of the distribution of effort or other laws of learning on the basis of results obtained from meaningless materials and then assume that they are equally true for meaningful materials.

APPLICATION TO THE LEARNING OF SCHOOL SUBJECTS

If meaningful prose is easily learned and well remembered, it is worth while investigating to what extent this type of learning can be improved, for evidently the learning and retention of meaningful material is about the most important task of students and teachers. A number of investigations have been made to show that the learning and retention of prose substance can be greatly improved. Some of the devices which experimenters have found to be effective are: Giving advance questions on the main points of the selection to be studied, placing leading questions at the head of the paragraphs, simplifying the vocabulary, giving advance directions requiring specific responses, giving advance explanations of content, reading with the intention of finding and retaining the central thought, sentence completion exercises based on the thought of the story, finding the relationship of the title to the central thought, and assigning definite problems to be solved by reading.[29] Numerous experiments could be cited to prove the effectiveness of these devices, but lack of space prevents us from describing them. One reason for the helpfulness of these devices is that they create a mental set which is adapted to the task in question. This applies to all of the devices that call for a more or less definite response from the reader, after the readings are completed. The helpfulness of a simplified vocabulary is due to another factor; namely,

the ability of simple words to arouse meaningful associations. The foregoing devices apply to any meaningful prose regardless of the subject-matter with which it deals.

The factor of meaning is, however, not limited to the study of meaningful prose. It has applications to all school subjects, even in the learning of combinations in arithmetic, the learning of spelling, and the learning of typewriting, which are ordinarily thought of as learned principally by drill devices. An interesting experiment which shows the importance of emphasizing relationships in the teaching of combinations in arithmetic was made by Thiele[32] in the public schools of Detroit. Instead of presenting number combinations in their order of difficulty, as is the custom, giving one at a time, and as often as is necessary until it is completely learned, he presented the combinations in such a manner as to emphasize relationships. For example, he presented the ten digits in a row, adding one to each digit, then two to each digit, then three to each digit, etc. This enabled the learner to see readily the effect of adding the given number to each digit. It made the sum larger by just this number. The per cent of error in this mode of presentation was from four to thirteen per cent less than from the traditional mode. Another investigator that took advantage of the factor of meaning to learn number combinations was Wilson. Instead of teaching combinations by the drill method, he had first-grade children learn them from games in which numerical scores were kept. At the end of a year, the subjects used in the experiment had accuracy scores ranging from ninety-seven to one hundred, instead of the usual sixty-five per cent obtained in most standardized tests.

The introduction of meaning is useful not only in teaching the elementary combinations, but also in problem solving. Here one of the most important factors has been found to be the familiarity of the situations and terms of the problem. Other helpful devices are emphasis on generalizations or general ideas of procedure, the use of systematic analysis, and the use of diagraming rather than the centering of attention upon getting the right answers. The helpfulness of these[29] devices is probably due to the fact that they enable pupils to perceive the significant relationships involved in a problem.

The traditional mode of teaching spelling is the drill method, repeating the letters over and over until they can be reproduced without error. Everyone who has tried to reason out how a word should be spelled has found the result to be very annoying. In spite of this fact, experiments show that defining the meaning of the words

taught adds very greatly to the accuracy of the spelling.[29] In fact, this has been one of the most helpful devices that has so far been discovered in the economical teaching of spelling. Typewriting is another subject which is learned principally by repetition. The traditional method of teaching the subject is to spend the first half of a semester in giving drills on meaningless material. An experiment made by Barton[4] of the University of Idaho showed that typewriting is learned very much more rapidly if meaningful material is used instead of meaningless material. He had his subjects write a real business letter on the first day, using a chart on the board as a guide for striking the correct keys. His results showed that in proportion to the amount of time spent in learning, the pupils who had connected material made four and a half times as much gain as the pupils who used disconnected material.

Since most of the material used in the instruction of teaching social studies and science is reading material, it is reasonable to suppose that those devices used for increasing comprehension of meaningful prose can be applied directly to the reading of materials in these subjects. Experiments have shown that this is the case. Two devices that have been found to be particularly valuable in teaching social science are outlining[6] and simplification of the vocabulary.[3] Both increase meaning.

The study-guide method has been found to be very profitable in teaching high-school zoology[24] and physics.[7,9] Its aim is to make the student capable of independent study and research, to train the student how to think in a given subject rather than to acquire information about it, to give the student an idea of the kind of questions that are asked in a given subject, and how to find the answers rather than merely to train him to master the facts about it. Two important characteristics of the method are diminishing guidance and the so-called learning threads. In diminishing guidance, the plan calls for much help when the student starts a new subject, but, after this, a reduction in the amount of help as rapidly as possible, until the student can get along without any. Learning threads refer to the questions that are asked about every topic in a given subject. For example, in zoology, we want to know certain things about every specimen, such as: How is it classified? What is its structure? What is its habitat? What is its food? How does it reproduce? What is its economic value? etc. When a student knows what questions to ask and how to find the answers, he can take up one topic after another, and secure a

systematic knowledge abou it. When he has reached this point, he is capable of independent study, which is the result desired. Just as there are certain questions asked about a biological specimen, so there are characteristic questions asked about every field of study, and they are the learning threads which enable the pupil to transfer his technique of investigation from one topic to another. Experiments show that high-school students make considerably more improvement in physics and zoology when they follow the study-guide procedure than when they use the traditional procedure. These enlarged gains are undoubtedly due to the fact that the study-guide method emphasizes the importance of meaning, or significant relationships.

MEANING AND THEORY OF LEARNING

The studies made of meaning as a factor in learning have already produced important changes in our conceptions of the theory of learning. The principles of contiguity, practice, the conditioned reflex, and the law of effect have receded from the foreground; and in their place have come such terms as insight, understanding, intelligence, organization, pattern, and the like.[20,21,34] There is much mystery in the content of these terms, but when more precisely defined they will lead us to not only a truer understanding of the learning process, but also to more practical results.

The foregoing studies indicate that meaning must be recognized as a major factor in learning, that it has possibilities of wide application, and that as a topic for research it presents a field which is not only comparatively new and untouched, but also promising of fruitful results.

BIBLIOGRAPHY

1. Anastasi, A.: "Further Studies on the Memory Factor." *Archives of Psychology*, Vol. XXII, 1932–1933, 60 pp.
2. Austin, S. D. M.: "A Study of Logical Memory." *American Journal of Psychology*, Vol. XXXII, 1921, pp. 370–403.
3. Ayer, Adelaide, M.: *Some Difficulties in Elementary School History.* Contributions to Education, No. 212, Teachers College, Columbia University, New York, 1926.
4. Barton, J. W.: "Comprehensive Units in Learning Typewriting." *Psychological Review Monographs*, Vol. XXXV, 1926, No. 164.
5. Barton, Wm. A.: *Outlining as a Study Procedure.* Contributions to Education, No. 411, Teachers College, Columbia University, New York, 1930, 115 pp.
6. Bennett, F.: "Correlations Between Different Memories." *Journal of Experimental Psychology*, Vol. I, 1916, pp. 404–418.

7. Blank, Irene: *An Experiment in Supervised Instruction in Physics.* Master's Thesis, University of Pittsburgh, 1928.
8. Bryan, W. L., and Harter, N.: "Studies in the Physiology and Psychology of the Telegraphic Language." *Psychological Review*, Vol. IV, 1897, pp. 27–53, Vol. VI, 1899, pp. 345–375.
9. Clemensen, Jessie W.: *Study Outlines in Physics.* Contributions to Education, No. 553, Teachers College, Columbia University, New York, 1933.
10. Dietze, A. G., and Jones, G. E.: "Factual Memory of Secondary School Pupils for a Short Article Which They Read a Single Time." *Journal of Educational Psychology*, Vol. XXII, 1931, pp. 586–598, 667–676.
11. Ebbinghaus, H.: *Memory.* Translated by H. A. Ruger, Teachers College Columbia University, 1913, pp. VIII + 123.
12. English, H. B., Welborn, E. L., and Killian, C. D.: "Studies in Substance Memorization." *Journal of General Psychology*, Vol. XI, 1934, pp. 233–260.
13. English, H. B., and Welborn, E. L.: "Logical Learning: A General Review of Experiments with Meaningful Verbal Materials." *Psychological Bulletin*, Vol. XXXIV, 1937, pp. 1–20.
14. Foster, J. C.: "Verbal Memory in Pre-School Children." *Journal of Genetic Psychology*, Vol. XXXV, 1928, pp. 26–44.
15. Gates, A. I.: "Recitation as a Factor in Memorizing." *Archives of Psychology*, No. 40, 1917.
16. Hartmann, Geo. W.: *Gestalt Psychology.* Ronald Press Company, New York, 1935, pp. XIII + 325.
17. Hunsley, Yuba: *The Influence of the Number of Readings on the Rate of Forgetting.* Master's Thesis, Fort Hays Kansas State College, 1933.
18. Jones, M. G., and English, H. B.: "Notional *vs.* Rote Memory." *American Journal of Psychology*, Vol. XXXVII, 1926, pp. 602–603.
19. Judd, C. H.: "The Relation of Special Training to Intelligence." *Educational Review*, Vol. XXXVI, 1908, pp. 28–42.
20. Köhler, Wolfgang: *Gestalt Psychology.* Horace Liveright, New York, 1929, pp. XI + 403.
21. Koffka, Kurt: *The Principles of Gestalt Psychology.* Harcourt Brace, and Company, 1935, pp. XI + 724.
22. Lyon, D. Oliver: "The Relation of Length of Material to Time Taken for Learning; the Optimum Distribution of Time." *Journal of Educational Psychology*, Vol. V, 1914, pp. 1–9, 85–91, 155–163.
23. Peterson, J. C.: "The Higher Mental Processes in Learning." *Psychological Review Monographs*, Vol. XXVIII, 1920, No. 129.
24. Predmore, Donald, R.: *An Experiment in Supervised Instruction in Zoology.* Master's Thesis, University of Pittsburgh, 1930.
25. Pyle, W. H.: *The Examination of School Children.* The Macmillan Company, New York, 1913.
26. Reed, H. B.: "Associative Aids I. Their Relation to Learning, Retention, and Other Associations." *Psychological Review*, Vol. XXV, 1918, pp. 128–155.
27. ———: "Associative Aids II. Their Relation to Practice and the Transfer of Training." *Psychological Review*, Vol. XXX, 1918, pp. 257–285.
28. ———: "Repetition and Association in Learning." *Pedagogical Seminary*, Vol. XXXI, 1924, pp. 147–155.

29. ———: *Psychology of Elementary School Subjects.* Ginn and Company, Boston, 1938., pp. XI + 582.
30. Ruger, H. A.: "The Psychology of Efficiency." *Archives of Psychology,* No. 15, 1910.
31. Shaw, J. C.: "A Test of Memory in School Children." *Pedagogical Seminary,* Vol. IV, 1896, pp. 61–78.
32. Thiele, C. L.: "The Mathematical Viewpoint Applied to the Teaching of Elementary Arithmetic." *Tenth Yearbook of the National Council of Teachers of Mathematics,* 1935, pp. 212–233.
33. Thorndike, E. L.: "Animal Intelligence: An Experimental Study of the Association Processes in Animals." *Psychological Review Monographs,* Supplement II, 1898, No. 4, pp. 1–109.
34. Wheeler, R. H., and Perkins, F. T.: *Principles of Mental Development.* Thomas Y. Crowell Co., New York, 1932, pp. XIII + 529.

SYSTEMS OF REGRESSION EQUATIONS FOR THE ESTIMATION OF FACTORS

HARRY H. HARMAN

University of Chicago

1. METHODS OF APPRAISING FACTOR ABILITIES

Unless we can deduce some practical and reliable means of appraising an individual's factor abilities from a given factor solution, the entire problem of factor analysis would seem to be fruitless. Several[1] proposals for the solution of this problem have been made, but none with useful consequences. The trouble with most of these methods is that, although they give the most valid results statistically, the time element makes them prohibitive. It is the purpose of this paper to analyze some of the existing methods of appraising factors and to arrive at a method which shall be practical and statistically sound.

We recognize five distinct methods of appraising abilities which will be described in the following paragraphs.

I. Complete Estimation.—By this method we mean the estimate of any factor by means of a regression equation involving all the tests. This method has been proposed by Thurstone[2] and Thomson.[2]

II. Two-factor Estimation.—This method is a variation of the original method set forth by Spearman to meet more complex situations.[3] The general procedure is to reduce a complex factor solution to a two-factor form by combining the tests that measure group factors, and then to estimate each factor directly from the composite test that measures it. This is always possible with the bi-factor solution (in which there is no overlapping between group factors), but may be impossible with a complex factor solution in which there is overlapping

---

[1] Thurstone, L. L.: *The Vectors of Mind.* Chicago: The University of Chicago Press, 1935, Chap. X.

Thomson, G. H.: "Some Points of Mathematical Technique in the Factorial Analysis of Ability." *The Journal of Educational Psychology*, January, 1936, pp. 37–55.

Holzinger, K. J.: *Statistical Resumé of the Spearman Two-Factor Theory.* Chicago: The University of Chicago Press, 1930.

Preliminary Report on Spearman-Holzinger Unitary Trait Study, No. 8. Prepared at the Statistical Laboratory, Department of Education, The University of Chicago, 1936.

[2] *Op. cit.*

[3] See Holzinger: *Statistical Resume of the Spearman's Two Factor Theory.*

between group factors. Holzinger has described and applied this method to actual data.[1]

III. Reduced Estimation.—This method is a combination of Methods I and II. The complete estimation method is applied to the composite tests, as a means of shortening the work of Method I.

IV. Partial Estimation.—The regression equations in this method involve only the tests which best measure or define the factor under consideration in each case.

V. Complete Approximation Estimation.—We propose this method, which is a combination of Methods III and IV, as the best statistical approximation to the complete estimation method while requiring much less time. The procedure is to combine the respective tests that measure the factors other than the one under consideration, and to employ these composite tests along with the tests measuring this factor in the regression equation.

2. ILLUSTRATION OF METHODS

The preceding methods of estimating factors will be illustrated on an hypothetical problem of twelve tests for which we have a number of factor solutions.[2] In Table I are presented the intercorrelations of the twelve tests.

TABLE I.—INTERCORRELATIONS OF TWELVE TESTS

	1	2	3	4	5	6	7	8	9	10	11	12
1												
2	.15											
3	.20	.20										
4	.25	.25	.33									
5	.11	.11	.16	.19								
6	.13	.13	.18	.22	.56							
7	.10	.10	.13	.16	.41	.48						
8	.13	.13	.18	.22	.56	.66	.48					
9	.06	.06	.08	.10	.08	.09	.06	.09				
10	.09	.09	.12	.15	.12	.14	.10	.14	.10			
11	.12	.12	.16	.19	.16	.18	.13	.18	.13	.20		
12	.15	.15	.20	.24	.20	.23	.17	.23	.17	.26	.34	

[1] *Preliminary Report on Spearman-Holzinger Unitary Trait Study*, No. 8.

[2] This problem was set up by Holzinger and Harman in "Relationships between Factors Obtained from Certain Analyses." *The Journal of Educational Psychology*, May, 1937, pp. 321–346.

In order to appraise a set of factors it is necessary to know the factor structure[1] of the battery of tests. We shall use the structures obtained by Holzinger and the writer for the following solutions: (1) *Bi-factor*, a system of independent factors such that each test is expressible as a linear function of a general and a group factor; (2) *multiple orthogonal*, a solution in which each test is expressed in terms of a number of independent factors; (3) *primary*, a system of correlated factors obtained by referring the tests to the primary axes, *i.e.*, the lines of intersection of $(N - 1)$ of the N reference hyperplanes.[2]

For each factor solution we shall obtain the different factor estimates and compare their multiple correlation coefficients as a basis for judging the statistical soundness of the methods. The complete estimation method shall be used as the standard, since it produces the highest multiple correlation, and we shall seek to approximate this method by another which is more parsimonious.

Before proceeding to the analysis, may we comment on the form of computation. After trying various methods for the solution of regression equations, we have come to the following conclusions: The determinant method is very convenient for three or four variables.[3] A variation of the determinant method (which makes use of a short-cut method of expanding determinants) is convenient in the case of three to six variables.[4] Finally, the Doolittle method is by far the most practical for more than four variables.[5]

The correlations of the twelve tests with the different factors are presented in the factor structures of Table II.

From the intercorrelations of the tests, given in Table I, and the correlations of the tests with the factors, we obtain the complete estimation for each factor. These equations of estimation are pre-

[1] By a factor structure we mean a table of correlations of the tests with the factors. In contrast to this, by a factor pattern we mean a set of equations expressing the tests as linear functions of the factors, where all variables are in standa d form. It can easily be shown that the elements of a factor structure become the coefficients in the factor pattern if the factors under consideration are statistically independent.

[2] Thurstone: *Op. cit.*, Chapter VI.

[3] Holzinger, K. J.: *Statistical Methods for Students in Education.* Boston: Ginn and Company, 1928, pp. 312–316.

[4] A description of this method, which is ascribed to Aitken, may be found in Thomson: *Op. cit.*

[5] See Peters, Charles C. and Wykes, Elizabeth Crossley: "Simplified Methods for Computing Regression Coefficients and Partial and Multiple Correlations." *Journal of Educational Research*, Vol. XXIII, 1931, pp. 383–393.

TABLE II.—FACTOR STRUCTURES

Test j	Bi-factor				Multiple orthogonal			Primary		
	r_{ju}	$r_{j\alpha}$	$r_{j\beta}$	$r_{j\gamma}$	r_{jZ_1}	r_{jZ_2}	r_{jZ_3}	r_{jT_1}	r_{jT_2}	r_{jT_3}
1	.29	.26	...	...	.36	.13	.02	.38	.16	.23
2	.29	.26	...	...	.36	.13	.02	.38	.16	.23
3	.40	.32	...	...	.48	.18	.03	.51	.22	.31
4	.48	.43	...	...	.60	.23	.03	.64	.28	.38
5	.39	...	.56	...	.06	.69	.10	.30	.69	.31
6	.45	...	.68	...	.06	.80	.12	.35	.81	.36
7	.33	...	.49	...	.05	.58	.08	.25	.59	.26
8	.45	...	.68	...	.06	.80	.12	.35	.81	.36
9	.20	...	...	.16	.12	.07	.22	.15	.11	.26
10	.31	...	...	.24	.19	.11	.33	.23	.17	.39
11	.40	...	...	.33	.24	.15	.43	.30	.23	.52
12	.51	...	...	.43	.31	.18	.55	.38	.28	.66

TABLE III.—BI-FACTOR ESTIMATES

Method		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_A	x_B	x_C
I	$\hat{u}$	.09	.09	.15	.21	.07	.12	.05	.12	.06	.10	.15	.24			
	$\bar{\alpha}$	.15	.15	.20	.36	−.03	−.05	−.02	−.05	−.03	−.04	−.07	−.10			
	$\bar{\beta}$	−.04	−.04	−.08	−.12	.18	.37	.14	.37	−.03	−.05	−.08	−.13			
	$\bar{\gamma}$	−.04	−.04	−.06	−.09	−.03	−.05	−.02	−.05	.08	.14	.23	.38			
II	$\bar{\alpha}$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.488		
	$\bar{\beta}$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.751	
	$\bar{\gamma}$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.459
III	$\hat{u}$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.36	.30	.36
	$\bar{\alpha}$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.57	−.12	−.15
	$\bar{\beta}$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	−.19	.86	−.18
	$\bar{\gamma}$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	−.14	−.11	.53
IV	$\hat{u}$	Same as under Method I.														
	$\bar{\alpha}$	.13	.13	.17	.31											
	$\bar{\beta}$	.....	.....	.....	.....	.15	.33	.12	.33							
	$\bar{\gamma}$	.....	.....	.....	.....	.....	.....	.....	.....	.07	.11	.19	.33			
V	$\hat{u}_A$	.09	.09	.16	.22	.....	.....	.....	.....	.....	.....	.....	.....	.....	.29	.36
	$\hat{u}_B$	.....	.....	.....	.....	.07	.12	.05	.12	.....	.....	.....	.....	.36	.....	.36
	$\hat{u}_C$	.....	.....	.....	.....	.....	.....	.....	.....	.06	.10	.15	.25	.35	.29	
	$\bar{\alpha}$	.15	.15	.20	.36	.....	.....	.....	.....	.....	.....	.....	.....	.....	−.13	−.16
	$\bar{\beta}$	.....	.....	.....	.....	.18	.37	.13	.37	.....	.....	.....	.....	−.19	.....	−.19
	$\bar{\gamma}$	.....	.....	.....	.....	.....	.....	.....	.....	.08	.13	.22	.38	−.15	−.13	

sented in tabular form under Method I in Tables III, IV, and V. The dashed letters designate the factors being appraised. Each row represents an equation; the coefficients are given explicitly and the variable to which each is attached appears at the head of its column.

TABLE IV.—MULTIPLE ORTHOGONAL FACTOR ESTIMATES

Method		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_A	x_B	x_C
I	$\bar{Z}_1$	.17	.17	.27	.42	−.04	−.08	−.02	−.08	.02	.05	.06	.12			
	$\bar{Z}_2$	.00	.00	.00	.02	.23	.37	.14	.37	−.01	−.02	−.02	−.05			
	$\bar{Z}_3$	−.05	−.05	−.07	−.12	−.01	−.00	−.01	−.00	.11	.18	.27	.45			
III	$\bar{Z}_1$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.69	−.17	.17
	$\bar{Z}_2$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.02	.91	−.06
	$\bar{Z}_3$	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	−.17	−.01	.66
IV	$\bar{Z}_1$	.17	.17	.27	.42											
	$\bar{Z}_2$	.....	.....	.....	.....	.22	.37	.14	.37							
	$\bar{Z}_3$	.....	.....	.....	.....	.....	.....	.....	.....	.10	.17	.25	.41			
V	$\bar{Z}_1$	.17	.17	.27	.43	.....	.....	.....	.....	.....	.....	.....	.....	.....	−.17	.15
	$\bar{Z}_2$	.....	.....	.....	.....	.23	.37	.14	.37	.....	.....	.....	.....	.02	.....	−.06
	$\bar{Z}_3$	.....	.....	.....	.....	.....	.....	.....	.....	.11	.18	.27	.45	−.18	−.03	

We proceed now to the other methods of estimating factors. The first step is to reduce the factor structures by means of composite tests. It is evident from Table II that the groups of tests (1,2,3,4), (5,6,7,8),

TABLE V.—PRIMARY FACTOR ESTIMATES

Method		x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_A	x_B	x_C
I	$\bar{T}_1$	.16	.16	.25	.39	.03	.06	.02	.06	.02	.04	.07	.11			
	$\bar{T}_2$	.01	.01	.01	.04	.20	.36	.14	.36	.01	.01	.03	.02			
	$\bar{T}_3$	.04	.04	.07	.11	.03	.06	.02	.06	.10	.16	.26	.42			
III	$\bar{T}_1$	...	...	...	...	...	...	...	...	...	...	...	...	.63	.16	.17
	$\bar{T}_2$	...	...	...	...	...	...	...	...	...	...	...	...	.05	.88	.05
	$\bar{T}_3$	...	...	...	...	...	...	...	...	...	...	...	...	.21	.17	.61
IV	$\bar{T}_1$	.18	.18	.29	.45											
	$\bar{T}_2$	...	...	...	...	.21	.38	.14	.38							
	$\bar{T}_3$	...	...	...	...	...	...	...	...	.12	.19	.30	.49			
V	$\bar{T}_1$	.16	.16	.25	.40	...	...	...	...	...	...	...	...	...	.15	.16
	$\bar{T}_2$	...	...	...	...	.20	.37	.14	.37	...	...	...	...	.05	...	.04
	$\bar{T}_3$	...	...	...	...	...	...	...	...	.10	.16	.26	.42	.18	.15	

and (9,10,11,12) measure the respective group factors of the different solutions. Let us denote the standardized form of the sums of these tests as follows:

$$x_A = \frac{x_1 + x_2 + x_3 + x_4}{\sigma_{(1+2+3+4)}},$$
$$x_B = \frac{x_5 + x_6 + x_7 + x_8}{\sigma_{(5+6+7+8)}}, \tag{1}$$
$$x_C = \frac{x_9 + x_{10} + x_{11} + x_{12}}{\sigma_{(9+10+11+12)}}.$$

The standard deviation of a sum of n standardized variables is given by

$$\sigma^2_{(1+2+\cdots+n)} = n + 2 \sum_{i<j=1}^{n} r_{ij}. \tag{2}$$

The correlation between two standardized variables,

$$x_M = \frac{x_1 + x_2 + \cdots + x_m}{\sigma_{(1+2+\cdots+m)}},$$
$$x_N = \frac{x_1 + x_2 + \cdots + x_n}{\sigma_{(1+2+\cdots+n)}},$$

is given by the well known formula,

$$r_{MN} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} r_{ij}}{\sigma_{(1+2+\cdots+m)}\sigma_{(1+2+\cdots+n)}}. \tag{3}$$

The correlations between the composite tests A, B, and C may be computed by means of formula (3). They are presented in Table VI.

TABLE VI.—INTERCORRELATIONS OF COMPOSITE TESTS

	A	*B*	*C*
A			
B	.285		
C	.316	.283	

We can obtain the correlations of the composite tests with the different factors by use of formula (3). In this case the variable M

is one of the factors and the variable N is one of the composite tests (1). This yields the reduced factor structures of Table VII.

TABLE VII.—REDUCED FACTOR STRUCTURES

Composite test k	Bi-factor				Multiple orthogonal			Primary		
	r_{ku}	$r_{k\alpha}$	$r_{k\beta}$	$r_{k\gamma}$	r_{kZ_1}	r_{kZ_2}	r_{kZ_3}	r_{kT_1}	r_{kT_2}	r_{kT_3}
A	.562	.488	....	....	.692	.258	.038	.735	.315	.442
B	.505	....	.751	....	.072	.894	.131	.389	.904	.402
C	.561	....	....	.459	.340	.202	.605	.419	.312	.723

From the reduced bi-factor structure we read off directly the two-factor estimates of α, β, and γ. These appraisals appear under Method II in Table III. Since the general factor u is measured by all the tests we may obtain a two-factor estimate of it by using the composite test

$$x_u = \frac{x_A + x_B + x_C}{\sigma_{(A+B+C)}}.$$

Then in the complete reduction of the bi-factor structure, $r_{uz_u} = .746$. The estimate of u then becomes

$$\bar{u} = .746x_u. \tag{4}$$

Next, we obtain the reduced estimates from the intercorrelations of the composite tests and the correlations of these tests with the factors, given in the reduced structures. These estimates are presented under Method III in Tables III, IV, and V.

The partial estimates are simply obtained from the groups of tests which best measure the different factors. These appraisals are given as Method IV.

Finally, we pass to the complete approximation method of estimating these factors. This method makes use of the individual tests which best measure the factor in each case along with the remaining two of the three composite tests A, B, C. In order to obtain these estimates it is then necessary to have the correlations of the composite tests with the original twelve tests. These correlations may be computed by means of formula (3), and are presented in Table VIII. From the intercorrelations of the tests (given in Tables I, VI, and VIII) and the correlations of the tests with the factors (given in Tables

TABLE VIII.—CORRELATIONS OF COMPOSITE TESTS WITH ORIGINAL TESTS

	1	2	3	4	5	6	7	8	9	10	11	12
A	...	...	...	...	.22	.25	.19	.25	.12	.17	.23	.28
B	.15	.15	.20	.25	...	...	...	...	.10	.16	.20	.26
C	.17	.17	.22	.27	.22	.25	.18	.25				

II and VII), we obtain the complete approximation appraisals as presented under Method V. In applying Method V to the general factor of the bi-factor solution we have three alternatives for selecting the tests which individually measure u. We have denoted the estimates by $\bar{u}_A$, $\bar{u}_B$, $\bar{u}_C$ when using the tests (1, 2, 3, 4, *B*, *C*), (5, 6, 7, 8, *A*, *C*), and (9, 10, 11, 12, *A*, *B*), respectively.

Next, we shall obtain the multiple correlation coefficients of the estimates in Table III, IV, and V. In general,

$$R^2_{\tau\bar{\tau}} = \sum_{j=1}^{n} a_{\tau j} r_{\tau j}, \tag{5}$$

where

τ = any factor,
j = any test used in estimating τ,
n = number of tests j used in estimating τ,
$a_{\tau j}$ = coefficient of test j in the equation of estimation of τ,
$r_{\tau j}$ = correlation of test j with factor τ.

Applying formula (5) to the different factors, with j running over the respective tests used in estimating the factor in each case, we obtain the multiple correlation coefficients of Table IX. The following notation has been used in Table IX:

τ = any factor;
k = *A*, *B*, or *C*;
(i) = (1, 2, 3, 4), (5, 6, 7, 8) or (9, 10, 11, 12);
(i, k) = (1, 2, 3, 4, *B*, *C*), (5, 6, 7, 8, *A*, *C*), or (9, 10, 11, 12, *A*, *B*).

We are now ready to compare the different methods for estimating the factors. Since the complete approximation method gives the best statistical results, we shall obtain the differences of the multiple correlation coefficients of each of the other methods from the corresponding correlation of Method I. These differences are presented in Table X.

TABLE IX.—MULTIPLE CORRELATION COEFFICIENTS

Factor	I	II	III,	IV	V
	$R_{T(1,2,\cdots,12)}$	R_{Tk}	$R_{T(A,B,C)}$	$R_{T(i)}$	$R_{T(i,k)}$
u	.768	.746	.746	.768	.753
α	.545	.488	.527	.505	.545
β	.820	.751	.804	.769	.817
γ	.534	.459	.493	.492	.529
Z_1	.745	....	.723	.710	.742
Z_2	.910	....	.898	.906	.908
Z_3	.662	....	.626	.641	.660
T_1	.794	....	.772	.757	.794
T_2	.918	....	.909	.914	.917
T_3	.800	....	.776	.765	.800

Inspection of Table X reveals the fact that the estimation of the different factors by the complete approximation method agrees very closely with the estimation by the complete regression method. Hence the complete approximation method is a scheme for appraising factors which is practical and statistically accurate. Furthermore, the systems of regression equations for orthogonal or oblique factors are

TABLE X.—DIFFERENCES IN MULTIPLE CORRELATION COEFFICIENTS

	I – II	I – III	I – IV	I – V
u	.022	.022	....	.015
α	.057	.018	.040	.000
β	.069	.016	.051	.003
γ	.075	.041	.042	.005
Z_1	....	.022	.035	.003
Z_2	....	.012	.004	.002
Z_3	....	.036	.021	.002
T_1	....	.022	.037	.000
T_2	....	.009	.004	.001
T_3	....	.024	.035	.000

of the same general form and are computed with equal simplicity. It is in the actual psychological problem of appraising factors for each individual of a given population that a discrimination between orthogonal and oblique factors may be desirable. For, if we wish to appraise a given factor out of a factorial solution involving, say, three factors, we may do so in the case of independent factors with the com-

plete assurance that we are estimating an individual's ability in only that one factor. In the case of dependent factors, however, we must appraise all three factors to get a true picture of an individual's ability in any one factor because of the intercorrelations between the factors.

3. ESTIMATES FROM COMPARISON OF FACTORS

In this section we shall obtain estimates of the multiple orthogonal and primary factors by means of the algebraic equations linking these factors with the bi-factors.[1] We shall show the agreement between the new estimates and the complete approximation appraisals in each case, and hence obtain an alternative way of estimating factors when several factorizations are made of the correlational matrix.

The relationships between the multiple orthogonal factors and the bi-factors are given in the following equations:

$$\begin{aligned} Z_1 &= .6175u + .7307\alpha - .3291\beta - .0001\gamma, \\ Z_2 &= .4521u + .0014\alpha + .8855\beta - .1178\gamma, \\ Z_3 &= .4309u - .4141\alpha - .1147\beta + .8032\gamma. \end{aligned} \tag{6}$$

Now, if we replace u, α, β, and γ in these equations by any appraisals of these factors we obtain estimations of the multiple orthogonal factors.

Substituting the complete approximation estimates $\bar{u}_A$ and $\bar{\alpha}$, and the reduced estimates $\bar{\beta}$ and $\bar{\gamma}$ into the first of equations (6) for u, α, β, and γ, respectively, we obtain the following estimate of Z_1:

$$Z_1' = .165x_1 + .165x_2 + .245x_3 + .399x_4 + .063x_A - .199x_B + .164x_C.$$

Now, if we replace x_A by its value $\dfrac{x_1 + x_2 + x_3 + x_4}{2.6000}$, this equation becomes

$$Z_1' = .19x_1 + .19x_2 + .27x_3 + .42x_4 - .20x_B + .16x_C. \tag{7}$$

Similarly, by substituting the complete approximation estimates of u and β, and the reduced estimates of α and γ into the second of equations (6) we get a new estimate of Z_2 which reduces to

$$Z_2' = .20x_5 + .39x_6 + .14x_7 + .39x_8 + .01x_A - .07x_C. \tag{8}$$

In like manner the estimate of Z_3 from the estimates of the bi-factors becomes

$$Z_3' = .12x_9 + .18x_{10} + .27x_{11} + .44x_{12} - .18x_A - .03x_B. \tag{9}$$

[1] These equations were obtained by Holzinger and Harman: *Op. cit.*, Section 9.

On comparing equations (7), (8), and (9) with the equations represented under Method V in Table IV we see that there are only slight discrepencies. Thus, the relationships (6), between multiple orthogonal factors and bi-factors, may be taken as the relationships between the estimates of these factors.

Finally we may obtain estimates of the primary factors from the following algebraic relationships between these factors and the bifactors:

$$\begin{aligned} T_1 &= .7587u + .6721\alpha + .0052\beta - .0133\gamma, \\ T_2 &= .5558u + .0018\alpha + .8340\beta - .0052\gamma, \\ T_3 &= .7821u + .0102\alpha + .0055\beta + .6261\gamma. \end{aligned} \tag{10}$$

Substituting the appropriate estimates of the bi-factors (as we have done in the preceding case) into (10), we obtain the following appraisals of the primary factors:

$$T_1' = .17x_1 + .17x_2 + .26x_3 + .41x_4 + .14x_B + .16x_C; \tag{11}$$
$$T_2' = .19x_5 + .37x_6 + .14x_7 + .37x_8 + .04x_A + .04x_C; \tag{12}$$
$$T_3' = .10x_9 + .16x_{10} + .25x_{11} + .43x_{12} + .18x_A + .15x_B. \tag{13}$$

These estimates are very nearly identical with the equations represented under Method V in Table V.

It has thus been shown that if appropriate estimates of the bi-factors are put into the equations linking the several systems of multiple-factors with the bi-factors, we obtain appraisals of the multi-factors that are very much like the complete approximation estimates in each case. Such estimates are statistically accurate, and hence the equations giving the relationships between two sets of factors may be considered as also giving the relationships between certain estimates of these factors. Thus, if two distinct factorial solutions are given, and the relationships between their factors are known, then the linkage between the appraisals of the two sets of factors is also known.

SCALE CALIBRATION BY THE THOMSON METHOD[1]

KATHERINE PRESTON BRADWAY[2]

The Training School at Vineland

Assigning age values to items of psychological tests has presented a problem ever since the Binet scale was first introduced. The problem raises the question: If the per cent of individuals passing an item at each of successive age levels is known, what is the average age of passing that item? Or, if an ogive type of curve shows the per cent passing at each successive age level, how can its central tendency be determined? The problem is a familiar one in the field of psychophysics where three methods are used for solving it. According to one method the fifty per cent point of the curve is determined either directly or by interpolating between the age level immediately below the fifty per cent point and the age level immediately above the fifty per cent point. For a second method a smooth curve is fitted to all the points of per cents according to a statistical formula and the central tendency of this curve is determined and used to represent the central tendency of the data. For a third procedure an arithmetic mean is calculated directly.

The first method was advocated by Thurstone for scaling items in 1925.[3] It has recently been used by Bayley.[4] The method is easy to apply, but it lacks the mathematical accuracy demanded in statistics since it ignores all the data except one or, at the most, two points. The result corresponds to the median and is subject to the same criticisms to which the median is subject.

Both the second and third methods take account of all points on the curve and provide, in addition to the central tendency, a measure of dispersion. The second method, that of fitting curves, was used by

---

[1] Presented at the annual meeting of the American Psychological Association at Minneapolis, Minnesota, September 2, 1937.

[2] The author wishes to acknowledge her indebtedness to Edgar A. Doll, Director of the Vineland Laboratory, for assistance in the several studies cited, and to George Kreezer, Research Associate at the Vineland Laboratory, for helpful suggestions in the writing of this paper.

[3] Thurstone, L. L.: "A Method of Scaling Psychological and Educational Tests." *Journal of Educational Psychology*, Vol. XVI, 1925, pp. 433–451.

[4] Bayley, N.: *Development of Motor Abilities During the First Three Years.* Monograph of Society for Research on Child Development, 1935, No. 1, p. 26.

Kreezer and Dallenbach[1] to determine the age at which children learned a certain relation. The procedure followed was identical with that used in the method of constant stimuli. This method involves concepts which are relatively unfamiliar in the field of psychometrics and is somewhat cumbersome to apply.

The third method, on the other hand, is based on a well-known concept and is relatively easy to apply. The determination of an arithmetical mean from a frequency distribution or curve is, of course, a familiar procedure. Since the ogive or per cent curve represents the summation of a frequency curve, it can readily be transformed into a frequency curve and the mean then determined. However, it is unnecessary to make this transformation if one uses the Hardy summation method of computing averages, since this method furnishes a procedure for calculating the mean directly from the per cent distribution. It is this method which Godfrey Thomson, in 1926, suggested be used for the calibration of items in psychological scales.[2] Despite its simplicity and statistical merit, the method has apparently received little attention.

Application of the method requires that the per cent distribution curve be complete, or nearly complete, with approximately zero per cent passing at the lowest age level and one hundred per cent passing at the highest age level, and that the age levels be equal units, as, for example, one year. The mean is then calculated by summing the percentages, dividing by one hundred and subtracting the resulting quotient from the upper limit of the highest level. The ogive curve is used to show the increase in the per cent of children passing a particular item of the scale. Suppose the percentages of children passing a particular item are as follows: Zero per cent at zero to one year, zero per cent at one to two years, twenty per cent at two to three years, sixty per cent at three to four years, ninety per cent at four to five years, and one hundred per cent at five to six years. The sum of these per cents is two hundred seventy. This sum divided by one hundred yields the quotient 2.7. The upper limit of the highest age level is 6.0, and 6.0 minus 2.7 gives 3.3, which is the average age for passing this particular item. A modification which eliminates the last step is to sum the per cent of failures from the zero to one year

[1] Kreezer, George and Dallenbach, Karl M.: "Learning the Relation of Opposition." *American Journal of Psychology*, Vol. XLI, 1929, pp. 432–441.

[2] Thomson, Godfrey, H.: "A Note on Scaling Tests." *Journal of Educational Psychology*, Vol. XVII, 1926, pp. 550–552.

level sand then forward mark off two places. In this illustration the sum of the failures equals three hundred thirty, and the average age would again be 3.3. A standard deviation is calculated by the formula:

$$SD = \sqrt{\frac{\Sigma \text{ acc per cent}}{100} - \frac{\Sigma \text{ per cent}}{100} - \left(\frac{\Sigma \text{ per cent}}{100}\right)^2}$$

where Σ acc per cent is the sum of the accumulated per cents, and Σ per cent is the sum of the per cents.

This Thomson method of calibrating items has been valuable in the preliminary standardization and validation of the Vineland Social Maturity Scale[1] which is a scale consisting of one hundred seventeen behavioral items. To demonstrate the use of the Thomson method, we shall describe the way it was used in each of five phases, and shall report a few sample results in order to illustrate the type of results which may be obtained.

The preliminary standardization of the Vineland Social Maturity Scale was based on its administration to six hundred twenty subjects.[2] This total included ten members of each sex at each age level from the zero to one year level through the thirty to thirty-one year level.

The first step in the standardization, and one which illustrates the first of our five uses of the Thomson method, was in the study of the value of individual scale items. Two of the factors which appeared important in order that an item be satisfactory were: (1) Absence of significant sex differences, and (2) Good differentiation of successive age levels in success in passing an item.

Sex differences on the items were determined by calculating average age values and standard deviations for both sexes according to the Thomson method. The difference between the age values for the two sexes divided by the standard error of the difference was then calculated. The conventional practice of accepting a critical ratio of over three as indicative of a significant difference was adopted in this study.

To study factor Number 2, the extent to which success and failure in passing items differentiated successive age levels, the mean age value

---

[1] Doll, Edgar A.: *The Vineland Social Maturity Scale: Revised, Condensed Manual of Directions.* Publication of the Training School at Vineland, New Jersey, Department of Research, Series 1936, No. 3, April, 1936, pp. 30.

[2] Doll, Edgar A.: "Preliminary Standardization of the Vineland Social Maturity Scale." *American Journal of Orthopsychiatry*, Vol. VI, April, 1936, 283–293.

and standard deviation for the sexes combined were calculated. The standard deviation (or SD) was then used as an indicator of the differentiating power of an item. This was done on the basis that a large SD represented a flat curve of per cents of passes at successive ages, and hence poor differentiation between the ages. Since inspection of the data and common assumption indicated that the SD increased with life age, the SD was related to life age by computing a coefficient of variability, *i.e.*, SD/Mean $\times$ 100. A coefficient of variability less than fifty was accepted as indicating satisfactory differentiating power.

Each item which showed a critical sex ratio greater than three, or a coefficient of variability greater than fifty, was modified or replaced by a new item. The six hundred twenty Social Maturity Scales administered for the standardization were then rescored according to the changes made in the items from the supplementary information which had been obtained in the course of administering the scales. None of the revised or substitute items showed a significant sex difference, or a coefficient of variability over fifty. (Additional criteria were used in selecting items, but a discussion of them is not pertinent to this paper.)

The second use which we made of the Thomson method was in the scaling of items. The items were placed in order of age value as determined by the Thomson method, and were separated into age levels according to norms determined on the basis of mean scores on the total test for successive age levels. This separation was done without regard for the age value of the items. For example, since a score of seventy-seven represented the score norm for 10.0 years, and eighty-one represented the score norm for 11.0 years, items 78 through 81 became the ten to eleven year items. For perfect item calibration, all items should have age values corresponding to the age-level group at which they appear. Thus items 78, 79, 80, and 81 should have age values between 10.0 and 11.0 years. Comparison of the age value for each item with the age at which the item was placed showed that there were only a few items which had age values above or below the age at which they had been placed. The standard errors of the age values of the items for which there was a discrepancy between age value and age of placement, approximated or were greater than the size of the discrepancy. Thus item 78, which, as it has been already pointed out, was at the ten-to-eleven year level, had an age value of

9.6, or .4 years below the lower limit of its year level. However, the standard error of this age value was .4 years. Consequently, no attempt was made to modify the items so as to change their difficulty.

The next step in the study of the scale was to validate each item against scores on the scale as a whole, *i.e.*, determine the extent to which success on each item was related to a high total score or social age. Scores on the Social Maturity Scale are expressed as social ages, just as Binet Intelligence Test scores are expressed as mental ages. Several regulation methods of item validation were tried, but none seemed satisfactory for the particular circumstances. However, the standard deviation of an item for successive social-age groups appeared to offer a satisfactory index of validity. Since it was desirable to have items which discriminated successive social-age groups, irrespective of life age, we used the Social Maturity Scale data for feeble-minded children whose life ages and social ages showed a negligible correlation, and in this way life age as a factor was controlled. The feeble-minded subjects were grouped according to social age, all those between social age zero to 1.0 in one group, between 1.0 to 2.0 in another, and so on. The per cent of each of the social-age groups passing each item was computed, and from these per cents the mean social-age value and standard deviation for each item of the scale for the feeble-minded were calculated. The mean age value represented the mean social age at which the feeble-minded pass a given item, and the standard deviation showed the dispersion of passes and failures in social-age groups above and below this mean social age. As an item approached perfect discrimination power between successive social ages, the dispersion of passes and failures became less, and the standard deviation approached zero. Consequently the smaller the standard deviation, the higher the validity of the item in terms of score on the total test.

The item standard deviations for the feeble-minded children grouped according to social age ranged from .4 to 3.4 for the sixty items of the scale for which we had complete enough data to calculate mean ages and standard deviations. Coefficients of variability were computed to allow for the increase of SD with social age. No item had a coefficient of variability greater than sixty.

This method of validating items against the scale as a whole takes into consideration all the data, and gives the result in easily understandable terms. A standard deviation of 1.0 for an item whose age value is 5.0 for example, may be interpreted as meaning that approxi-

mately seventeen per cent of subjects with a social age of 4.0 pass the item, and eighty-three per cent of subjects with a social age of 6.0 pass it. According to normal probabilities, we may predict that a subject who passed the item has a social age greater than 2.0 and a subject who fails it has a social age less than 8.0. It gives a method of validation which is comparable to other common statistical procedures and whose results are easy to interpret.

The fourth use of the results from the Thomson method was in determining which items of the scale had value in diagnosing feeblemindedness. The same data for the feeble-minded which were used for validation were used here. By comparing the social-age value of an item for the feeble-minded with the age value of the same item for the normal, it was possible to determine which items were "difficult" and which were "easy" for the feeble-minded. We determined the difference between the two age values for all items on the scale, and related this difference to the average of the two standard deviations. Because of the inequality of the number of cases at each social-age level, computation of the standard error of the means and hence standard errors of the differences, as was done for the two sexes in the standardization, was impracticable. Very few items appeared to have significant diagnostic value.

The fifth and last use to be mentioned was in the special handicapped groups. By using this method we could compare the difficulty of the items for the deaf[1] and the blind,[2] both with each other and with the hearing and the seeing. For this purpose subjects were grouped according to life age. The differences between the age values for each handicapped group and those for the normal were divided by the corresponding standard deviations for the normal group. If the differential ratio thus calculated was less than one, the difference was considered as not significant. If the ratio was one or over, the difference was considered as significant. It was found, for example, that the deaf were significantly retarded in sixty per cent of the items within the range of the scale used in the study, and the blind in ninety-three per cent.

We have indicated a few of the possibilities in using the Thomson method of scale calibration. We have used it for studying the value

[1] Bradway, Katherine P.: "The Social Competence of Deaf Children." *American Annals of the Deaf*, Vol. LXXXII, March, 1937, pp. 122–140.

[2] A report of the study of the social competence of the blind is now being prepared for publication.

of individual scale items, for scaling items, for validating items, for studying the diagnostic value of items, and for studying the specific handicaps of groups with sensory disabilities. The method probably has more ramifications. Its principal advantages are that it is based on well-known concepts, it is easy to apply, and the results lend themselves to further statistical treatment.

THE CONCEPT OF *INTELLIGENCE*

D. G. RYANS

Fulton, Missouri

It is appropriate to note that the introduction of the term "intelligence," with its present-day connotations, to the fields of psychological and educational thought and inquiry may be included in the intellectual chronology of the past half a century. True, the technical vocabulary of the epistemologist had long comprised the word, and it had been employed variously in the course of its evolution and translation from the Latin *intellectus* and the Greek cognate *nous*. Abstract in reference, there is little wonder, however, that no early serious attempt was made to apply its proposed meanings to the description of individual behavior.

Necessary to such a conception (*i.e.*, *intelligence* as a capacity and an attribute of the organism) was, first of all, an awareness of the universal existence of individual differences and, secondly, a recognition of the applicability of quantitative methods to the study of psychological phenomena. The transition occurred gradually, of course, and was accomplished, intentionally perhaps, through the efforts of Francis Galton, and, incidentally, as a result of acceptance of the Wundtian experiment and of the earlier dissemination of the theory of mental faculties by associationists. At the beginning of the last decade of the Nineteenth Century conditions were ripe for the induction of "intelligence" into the vocabulary and research of psychology and education. Gaining headway, falteringly but resolutely, interest reached its height following the World War—an interest which, if it seem less pronounced today, appears so only because its object has lost a bit of its novelty in becoming a part of the common stock and trade of contemporary learning.

Research and theory relative to "intelligence" have developed along many and heterogeneous lines, which, though easy to trace, present difficulties of organization. The effort to bring the materials together in an orderly statement is offered not without misgiving.

For convenience it may be well to consider preliminarily the growth of thought on *intelligence* with respect to three arbitrarily defined periods. The *first*, dating roughly from 1890 to 1910, appears to have been characterized chiefly by increasingly successful attempts to measure psychological phenomena and by the transition from the idea

of an "intelligence" revealed through sensori-motor processes to a belief in one involving, and identified by, highly complex performances. Little of basic importance seems to have been added to the theory of intelligence during a *second* period (1910–1920). This was, however, a most active phase during which widespread testing was executed, new tests and measure devised, and the results extensively applied. Inevitably excessive activity, while not purposively uncritical, often subordinates criticism to accomplishment. Such, in a limited sense, may be said to have been the case during the pre- and post-war years. During a *third* period, however, extending from the middle twenties to the present time, random testing has definitely diminished while an interest in analysis, refinement, and critique, both with respect to instruments and theories, has found expression.

It is, perhaps, seemly to review now the developing tendencies of thought on intelligence in slightly greater detail.

Note has already been taken of the prevailing attitude of the early nineties, which embraced a faith in the possibilities of tapping and sampling an individual's intellectual resources through the medium of his sensori-motor responses. Typically this point of view is attributable to Cattell who early popularized it, but there can be little doubt but that his inclinations were aroused and directed through contact with Francis Galton (whose position indicates pronounced influence from the associationists) and the practices in Galton's Antropometric Laboratories. Galton had written as early as 1883:

> The only information that reaches us concerning outward events appears to pass through the avenues of our senses; and the more perceptive the senses are of difference, the larger is the field upon which our judgment and intelligence can act.[1]

In accord with, and in defense of, this opinion that sensory discrimination provides the key to the determination of intelligence Cattell and his co-workers measured extensively such sensory, perceptual, and physical traits as could be made to yield to estimation.[2,3]

---

[1] Galton, F.: *Inquiries Into Human Faculty and Its Development.* London: MacMillan, 1883, p. 27.

[2] Cattell, J. McK.: "Mental Tests and Measurements." *Mind,* Vol. XV, 1890, pp. 373–380.

[3] Cattell, J. McK. and Farrand, L.: "Physical and Mental Measurements of the Students of Columbia University." *Psychol. Review,* Vol. III, 1896, pp. 618–648.

At the same time in Germany and France efforts were being made to attract attention toward strictly psychological tests (*i.e.*, measures of memory, association, judgment, attention, etc.) and a concept of intelligence which placed emphasis upon the more complex or highly organized "mental" activities. Oehrn[1] and Kraepelin[2] devised various test materials and expressed themselves on intelligence in terms of this view. Münsterberg[3] likewise suggested tests, involving relatively complex performances, but gave probably too much account to the time element and speed of response. Feeling that an organizing function, which was referred to as "*combination* ability," must supplement memory and other proposed mental traits in intelligent behavior, Ebbinghaus[4] as early as 1897 reported the "completion" test which not only undoubtedly influenced Binet's later work, but is still employed in many of our instruments.

However, Binet was already convinced of the futility of measuring intellect through the determination of special limited abilities. In collaboration with Henri in 1895 the comment was published:

> One cannot say, then, of an individual simply that he has a good memory; it is always necessary to state precisely of which memory one speaks. Therefore, in order to have an idea of an individual one must—include as many partial memories as possible.[5]

Recognizing the impossibility of testing all the dimensions in which individuals differ, Binet proposed to study a number of functions with regard to their respective differentiating capacities and to select those which appeared to be most characteristic and significant. Subsequent years found him investigating memory, imagination, suggestibility, attention, physical development, and even cephalic measurements in their relation to intelligence, and tentatively suggest-

---

[1] Oehrn, A.: "Experimentelle Studien zur Individualpsychologie." *Psychol. Arbeiten*, Vol. I, 1895, pp. 92–152.

[2] Kraepelin, Emil: "Der Psychologische Versuch in der Psychiatrie." *Psychol. Arbeiten.*, Vol. I, 1895, pp. 1–91.

[3] Münsterberg, H.: "Zur Individual Psychologie." *Centralblatt für Nervenheilkunde und Psychiatrie*, Vol. XIV, 1891, pp. 196–198.

[4] Ebbinghaus, H.: "Über eine neue Methode zur prüfung geistiger Fahigkeiten und ihre Anwendung bei Schulkindern." *Zeitschr. F. Angewandte Psychol.*, Vol. XIII, 1897, pp. 401–459.

[5] Binet, Alfred and Henri, V.: "La psychologie individuelle." *L'Année Psychologique*, Vol. II, 1895, pp. 411–465.

ing the comparison of individuals through the medium of graded instruments or scales.[1]

Early in his study of the subject Binet concluded that *attention* was a key to general ability. Attention was defined as "mental adaptation" to a novel situation. The paper entitled "Attention and Adaptation"[2] was published in 1900 dealing with the description and determination of voluntary attention. However, this field of endeavor was not, apparently, as yielding or profitable for his purposes as Binet had hoped. And a few years later, following a prolific period of research and writing, we observe him more firmly convinced than ever of the complexity of mental functions, but centering his descriptions of intelligence in such capacities as *direction and persistence of thought*, *autocriticism*, *invention*, and *comprehension*. In 1905 he wrote: "To judge well, to comprehend well, to reason well, these are the essential activities of intelligence."[3] Again in his *Les Idées Modernes sur les Enfants* an approximately similar view is expressed in more specific terms: "Comprehension, invention, direction, and criticism—intelligence is contained in these four words."[4]

The ultimate purpose of intelligence according to Binet is the continuous adjustment of the individual to his environment which is accomplished as the result of an organization and synthesis in which the several mental functions are involved. Thus, "the mind is unitary, despite the multiplicity of its faculties," and "intelligence . . . is, above all, a faculty of apprehension which is directed toward the external world and which works toward the reconstruction of it as a whole on the basis of the small fragments of it which we are given."[5]

It appears that Binet was not completely convinced of the hereditary nature of intelligence. In *Les Idées Modernes sur les Enfants* he states in no uncertain language that:

---

[1] Binet, Alfred: "La mesure en psychologie individuelle." *Revue Philosophique*, Vol. XLVI, 1898, pp. 113–123.

[2] Binet, Alfred: "Attention et Adaptation." *L'Année Psychologique*, Vol. VI, 1900, pp. 248–404.

[3] Binet, Alfred and Simon, T.: "Méthodes nouvelles pour le diagnostic du niveau intellectual des anormaux." *L'Année Psychologique*, Vol. XI, 1905, pp. 191–244.

[4] Binet, Alfred: *Les Idées Modernes sur les Enfants*. 1909, p. 118. (Quotations from Peterson's *Early Conceptions and Tests of Intelligence*.)

[5] *Ibid.*, p. 117.

With practice, ambition, and especially method one can succeed in increasing his attention, memory, and judgment, and in, literally, becoming more intelligent than before.[1]

Space does not permit discussion of Binet's scales devised for measuring mental ability. It is pertinent, however, to recall that the concept of *mental age*, while perhaps not wholly original with Binet, was first employed by him and first attained significance as the direct result of his applications. (For a treatment of Binet's all-important batteries of tests and their development, see such volumes as Peterson's *Early Conceptions and Tests of Intelligence*[2] and Pintner's *Intelligence Testing.*)[3]

In England during the early years of the Twentieth Century a view of the nature of intelligence which roughly approximated that of Binet was being expounded by Spearman. As early as 1904 he proposed:

All branches of intellectual activity have in common one fundamental function (or group of functions) whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others.[4]

The proof of his hypothesis Spearman believed to be demonstrable by the application of statistical techniques. Reliable coefficients of correlation between various test results were capable of arrangement so as to form a hierarchy when set up in tabular form. Furthermore, this hierarchy could be described in a series of definite mathematical formulae: *I.e., tetrad equations.* The conformance of correlation coefficients to these formulae indicated to Spearman the existence of two factors in each measurement, a "*g*" or general factor which might show itself in many measures and which would appear to be constant for the same individual, and an "*s*" or specific factor which was descriptive of a particular trait of the individual. Thus "*g*," which we may interpret as "general mental ability," would be common to all intellectual abilities, but would inclusively describe no one of them.

Spearman's professional interest throughout the past thirty-odd years has been centered in the accumulation of evidence to support his

---

[1] *Ibid.*, p. 143.

[2] Peterson, J.: *Early Conceptions and Tests of Intelligence.* Yonkers-on-Hudson, N. Y.: World Book Co., 1925.

[3] Pintner, R.: *Intelligence Testing*, New York: Henry Holt, and Co., 1931.

[4] Spearman, C.: "General Intelligence Objectively Determined and Measured." *Am. J. Psychol.*, Vol. XV, 1904, pp. 201–293.

two-factor theory of intelligence which has progressively developed during that time. In *The Nature of Intelligence and Principles of Cognition*[1] and *The Abilities of Man*[2] his position is treated in considerable detail. His basic concept of intelligence is found in his description of "*g*." It consists in the statement of a series of "qualitative" and "quantitative" principles (referred to as the *noegenetic laws*). The expression of "*g*" appears to consist of the mental activities involved in (1) *educing relations*, (2) *educing correlates*, and (3) *self-recognition*. These functions and the expression of "*g*" are, in turn, dependent on such conditions of the individual as (1) native limits, (2) retentivity, (3) fatigue, (4) effects of connation, and (5) physiological influences.

Binet and Spearman, it would seem, were both satisfied as to the existence of a general organizing factor of intelligence. Thorndike in this country, however, influenced by Cattell, took quite an opposite stand which denied the probability of a unitary concept of mind and suggested, rather, that intelligence, or intelligences, consisted of large numbers of relatively independent traits. Any correlation which might exist, among measures, said Thorndike, resulted, not from the existence of a common factor such as Spearman's "*g*" but, from the overlapping of functions which involved similar elements and common neural bonds. This position was first made known in 1902[3] extended in 1903[4] and reiterated many times during subsequent years. Today it is considered to be basically important to the theory of vocational psychology and personnel service.

One of Spearman's most severe critics has been a countryman, Thomson, who, by choosing playing-cards at random and correlating results, has attempted to show that chance samplings will contribute to a hierarchial matrix, and therefore that the fulfillment of the conditions of the tetrad equations by test data does not, in itself, prove the existence of a *general* intellectual factor.[5] Rather, says Thomson,

---

[1] Spearman, C.: *The Nature of Intelligence and Principles of Cognition.* London: MacMillan, 1923.

[2] Spearman, C.: *The Abilities of Man.* London: MacMillan, 1927.

[3] Aikins, H. A., Thorndike, E. L., and Hubbell, Eliz.: "Correlations among Perceptive and Associative Processes." *Psychol. Review*, Vol. IX, 1902, pp. 374–382.

[4] Thorndike, E. L.: *Educational Psychology.* New York: Lemcke and Buechner, 1903.

[5] Thomson, G. H.: "A Hierarchy without a General Factor." *Brit. J. Psychol.*, Vol. VIII, 1916, pp. 271–281. (Also other published papers.)

every psychological test samples a range of abilities and the extent of correlation between any two tests is determined by the number of abilities they have in common. This, of course, is in essence, the same view as that expressed by Thorndike in his description of independent abilities. Neither theory absolutely precludes Spearman's "*g*" (providing the sampling is extensive) but they both deny the necessity of it as an explanatory principle.

Psychologists and educators in the United States who became interested in the problem of intelligence, while influenced by the more immediate contacts of Cattell and Thorndike, rapidly accepted the Binet tests when they were introduced here. Goddard published translations of the 1905 and 1908 Binet-Simon scales, as did Whipple shortly after. Likewise, Kuhlmann introduced an American version along with comments on its administration.[1] The most widely employed of the American revisions of the Binet tests, however, has been that of Terman. Writing with reference to the use of the method as early as 1911,[2] Terman completed the well-known Stanford Revision and published it along with norms and an interpretation in 1916.[3]

The popular concept of intelligence today is closely associated with that of the Intelligence Quotient or IQ. It is fitting to recall, therefore, that this idea of measuring brightness or rate of mental growth, while prepared for by Binet in his use of mental-age levels, proposed by Stern[4] and employed by Bobertag as early as 1912[5] was introduced in this country by Terman and has become significant as the direct result of its association with the Stanford Revision of the Binet-Simon scale.

During the period of the World War at least two innovations in testing became known which, though technical rather than theoretical, undoubtedly influenced thinking on intelligence. These were, first of all, the testing of groups of assembled individuals (Pyle, Pintner, Otis, Miller) and secondly, the testing of persons who were, for one reason or another, incapable of understanding or responding to verbal

---

[1] Kuhlmann, F.: "Binet and Simon's System of Measuring the Intelligence of Children." *J. Psycho-Asthenics*, Vol. XV, 1911, pp. 76–92.

[2] Terman, L. M.: "The Binet-Simon scale for Measuring Intelligence: Impressions gained by its application." *Psychol. Clinic*, Vol. V, 1911, pp. 199–206.

[3] Terman, L. M.: *The Measurement of Intelligence.* Boston: Houghton Mifflin Co., 1916.

[4] Stern, W.: *The Psychological Methods of Testing Intelligence.* (Trs. Whipple, G. M.) Baltimore: Warwick and York, 1914.

[5] Bobertag, O.: "Über Intelligenz-prüfungen nach der Methode von Binet und Simon." *Zeitschr. fur angewandte Psychologie*, Vol. VI, 1912, pp. 495–538.

materials (Pintner and Paterson, *et al.*). Needless to say, also, the administration of the Army Examinations did much to encourage the development and analysis of measures of intelligence and, to an extent, to stimulate the clarification of thought and theory.

It is interesting to observe the opinion of investigators in the field of intelligence as revealed by a symposium conducted by the editors of the *Journal of Educational Psychology* and published in 1921.[1] That great diversity of opinion existed was but to be expected. Many of the views implied, in the first place, adherence to either one or the other of the fundamental standpoints set forth by Spearman and Thorndike. But they also included, for the most part, some description of the manifestations of intelligence. Thus, Thorndike referred to "intellect . . . as the *power of good response from the point of view of truth or fact.*"[2] Terman believed an individual to be intelligent "*in proportion as he is* able to carry on abstract thinking."[3] Pintner perpetuated the idea, to which both Binet and Stern had subscribed earlier, "of intelligence as the ability of the individual to adapt himself adequately to relatively new situations in life" implying "ease in breaking old habits and in forming new ones."[4]

"The capacity for knowledge and knowledge possessed"[5] constitute intelligence according to Henmon's view. Thurstone, at this date, referred to three phases of intelligent behavior, namely: "(*a*) Inhibitive capacity; (*b*) analytical capacity; and (*c*) perseverance."[6] Woodrow called intelligence "the capacity to acquire capacity."[7] "In terms of the current objective psychology," Dearborn defined intelligence as "the capacity to learn or profit by experience."[8] The trait was described by Haggerty as "a practical concept . . . connoting a group of complex mental processes traditionally defined in systematic psychologies as sensation, perception, association, memory, imagination, discrimination, judgment, and reasoning . . . emotions, instincts, will-activities, and so-called character traits" for the most part excluded.[9]

---

[1] Symposium: "Intelligence and its Measurement." *J. Educ. Psychol.*, 1921, pp. 123–147, 195–216.

[2] *Ibid.*, p. 124.

[3] *Ibid.*, p. 128.

[4] *Ibid.*, p. 139.

[5] *Ibid.*, p. 195.

[6] *Ibid.*, p. 202.

[7] *Ibid.*, p. 207.

[8] *Ibid.*, p. 210.

[9] *Ibid.*, pp. 212–213.

While not inclusive, the foregoing list indicates the more concise representative views of contributors to the symposium and would appear to reflect the thought of the period.

Several years later Thorndike, who had been constantly active in the measurement field, included "operations such as we may call attention, retention, recall, recognition, selective and rational thinking, abstraction, generalization, organization, inductive and deductive reasoning, together with knowledge and learning in general"[1] in his concept of intelligence, which he set about to describe more generally from the standpoint of its properties as revealed through tests.[2] Four different aspects of intelligence were considered, namely: *Altitude*, or level, which is apparently native and determines the limits of response with respect to difficulty; *range*, which may vary with experience, and which defines the extent of response at a given level; *area*, or totality of response, contributed to by both altitude and range; and *speed* of response. The descriptive value of Thorndike's classification has been widely recognized.

Another view with respect to intelligence, which has been developing during the past twenty years, is the opinion held by many of the Gestalt group that the quantitatively applied indirect methods employed by intelligence testers are not wholly acceptable, and that intellectual activity is revealed to a greater advantage through direct observation and qualitative description. Intelligent behavior according to this concept is defined in terms of a capacity[3] for means-end relationships. It consists in the perception "of one detail of a situation in a more complicated relationship to the total situation, namely, *the relationship of means to an end.*"[4]

Perhaps the most recent development in the growth of the concept of intelligence is the attention which is today being accorded the statistical analysis of test results in a search for group factors. We have often defined intelligence simply (and elusively, though with sincere intent) as "that which is measured by our so-called intelligence tests." Now developing out of Spearman's techniques devised for

---

[1] Thorndike, E. L., *et. al.: The Measurement of Intelligence.* New York: Teachers College, Columbia University, Press, 1927, p. 22.

[2] *Ibid.*

[3] Though employed here by the writer, *Gestalt* psychology does not condone the ideas of "capacities," "abilities," etc. on the ground that naive atomism is thus implied.

[4] Wheeler, R. H.: *The Science of Psychology.* New York: Thos. Y. Crowell & Co., 1929, p. 128.

the isolation and proof of a common intellectual factor, a method has been brought to the fore, which Thurstone, Kelley, and others believe will make possible the answering both of the question "what do the 'intelligence' tests measure?" and also that of "what is intelligence?"

Applying his own methods of factor analysis, Thurstone has announced the isolation of some seven primary factors which contribute to success in "intelligence test" performance. These may be conveniently referred to as (1) *number facility*, (2) *word fluency*, (3) *visualization* or *imagery*, (4) *memory* (not specific memories, but a general factor of memory), (5) *perceptual speed* or discrimination, (6) *induction* (generalization and abstraction), and (7) *verbal reasoning* or deduction.[1]

The emphasis upon group factors as the components of "intelligence" may be regarded, it seems, essentially as an intermediate position between the extremes suggested at the beginning of the century by Spearman and Thorndike and defended respectively by them throughout the years. This phase of research and theory is, of course, still in its infancy, but it would seem reasonable to anticipate its successful extension, and to propose that it is perhaps the most significant development of the present period and one upon which the future direction of energies in the field of individual psychology rests.[2]

[1] Thurstone, L. L.: "A New Conception of Intelligence." *Educ. Record*, Vol. XVII, 1936, pp. 441–450.

[2] Omissions in the above outline are conspicuous but necessary. The primary purpose has been not to include all contributions, but, to indicate the continuity of thought in the development of the concept of "intelligence."

A COMPARATIVE STUDY OF THE PSYCHONEUROTIC RESPONSES OF DEAF AND HEARING CHILDREN[1]

N. NORTON SPRINGER

Adolescents' Court of Brooklyn, New York

This article is part of an extensive study of personality, conducted in 1935–1936, on a group of New York City school children. Its purpose is to determine whether there are group differences between deaf and hearing children in their manifestations of maladjusted behavior as measured by their psychoneurotic responses on a personality inventory.

The Brown Personality Inventory for Children[2] was used for the measurement of psychoneurotic responses. This inventory consists of eighty questions, each of which is answered by *Yes* or *No*. These questions cover a wide variety of situations. The total number of atypical responses gives the neurotic score of the individual.

The Brown inventory was used because its language content appears to be fairly simple. Brown reports that the language of the inventory is not too difficult for the nine-year-old hearing child to understand.

The Goodenough Drawing of A Man Intelligence Test[3] was also given to the deaf and hearing children, in order to equate the groups and to study the effect of intelligence on psychoneurotic responses. The general social status of the children was determined by rating the parents' occupations on the Barr Scale of Occupational Status.[4]

Since the deaf child labors under a language handicap, the Brown inventory was given only to older children, who are able to read. Approximately four hundred deaf school children, with a mean age of sixteen years, attending three New York City public day and residential schools were used as the experimental group. The hearing

[1] The author wishes to acknowledge his indebtedness for assistance in this study to the Works Progress Administration Project on the Deaf, No. 165-97-8000, sponsored by Professor L. W. Max, of New York University.

[2] Brown, F.: "A Psychoneurotic Inventory for Children Between Nine and Fourteen Years of Age." *Journal of Applied Psychology*, Vol. XVIII, 1934, pp. 566–577.

[3] Goodenough, F.: "*Measurement of Intelligence by Drawings.*" Yonkers-on-Hudson, New York: World Book Co., 1926.

[4] Barr, F. E.: "Barr Social Rating Scale in Occupational Status," in Terman, L. M.: *Genetic Studies of Genius.* Vol. I, 2nd Ed., pp. 542–545, Stanford University Press, 1925.

control group was composed of three hundred twenty-seven children, in grades IV to IX, inclusive, attending the regular public schools. Difficulty was encountered in securing an adequate control group. The hearing children were about four years younger than the deaf children. This discrepancy in age, between the deaf and hearing groups, chosen from approximately the same school grades, is due to the retardation in the language development of the deaf. If the hearing children had been selected solely on the basis of chronological age, the control group would have consisted largely of academically retarded children. Had the deaf group been limited to pupils below the age of fourteen, very few cases would have been available. Since Brown reports that no relationship exists between the neurotic score and chronological age, intelligence, sex, and socio-economic status, the author decided to compare the older deaf and the younger hearing children. As a check on the comparison of older deaf children with younger hearing ones, the influence of chronological age on the neurotic scores was studied by means of the correlation technique.

Information in regard to chronological age, Barr Scale ratings, and Goodenough IQ's of the deaf and hearing groups, is given in Table I, in terms of the means, standard deviations, the mean differences between groups, the standard errors of these differences, and the critical ratios of the mean differences to their standard errors.

TABLE I.—CHRONOLOGICAL AGE, INTELLIGENCE, AND BARR SCALE RATINGS OF THE DEAF AND HEARING GROUPS

	Group	Deaf			Hearing			$M_D - M_H$	$SD_{diff.}$	Critical ratio
		Mean	SD	*N*	Mean	SD	*N*			
Age in years	Boys	16.32	2.44	225	12.78	2.11	190	3.54	.22	16.09
	Girls	16.22	2.47	172	11.23	1.59	137	4.99	.23	21.70
Good-enough IQ	Boys	88.01	16.08	209	87.80	19.64	168	.21	1.87	.11
	Girls	85.90	17.66	167	96.28	17.69	117	−10.38	2.11	4.92
Barr score	Boys	8.87	2.67	170	7.77	2.86	142	1.10	.47	2.34
	Girls	8.90	2.78	128	7.89	2.83	115	1.01	.82	1.23

Three hundred ninety-seven deaf and three hundred twenty-seven hearing boys and girls participated in this study. The mean age of the deaf groups was approximately sixteen years. The hearing chil-

dren's mean age was about four years less than that for the deaf group. These differences, as shown by the critical ratios, are statistically significant.

There is practically no difference in the mean IQ of the deaf and hearing boys. The mean IQ's for both groups of boys are eighty-eight. The hearing girls' mean IQ of ninety-six is ten points greater than the deaf girls' mean. This difference between the deaf and hearing girls is statistically significant. These IQ differences of the girls are similar to those reported by Springer[1] in a study of the intelligence of deaf and hearing children, between the ages of six and twelve. The mean IQ's for the older deaf and hearing children of this study are, however, about ten points lower than those reported for the six- to twelve-year-old groups. This may be due to the shortcomings of the Goodenough test in the measurement of the intelligence of older age groups.

The mean Barr Scale Ratings for the occupational status of the children's fathers indicate that there is a tendency for the large deaf groups to receive slightly higher ratings than those received by the hearing controls. These differences, however, are statistically unreliable, as indicated by the critical ratios of 1.23 and 2.34.

All the deaf and hearing children were born in New York City. More than ninety per cent of them are of foreign-born parentage.

1. RESULTS ON THE BROWN PERSONALITY INVENTORY

The results of the groups of deaf and hearing children are given in Table II. The mean neurotic score of the deaf boys, 37.43, is considerably higher than the mean score for the hearing boys, 21.03. The

TABLE II.—RESULTS OF THE DEAF AND HEARING CHILDREN ON THE BROWN PERSONALITY INVENTORY

Group	Deaf			Hearing			$M_D - M_H$	$SD_{diff.}$	Critical ratio
	Mean	SD	*N*	Mean	SD	*N*			
Boys	37.43	20.09	225	21.03	17.76	190	16.40	1.85	8.86
Girls	35.90	16.69	172	24.22	19.70	137	11.68	2.11	5.53
Total	36.75	18.70	397	22.40	18.65	327	14.35	1.39	10.32

[1] Springer, N. N.: "A Comparative Study of the Intelligence of a Group of Deaf and Hearing Children." *American Annals of the Deaf*, Vol. LXXXIII, 1938, pp. 138–152.

mean difference of 16.40 has a high degree of statistical reliability, as is indicated by the critical ratio of 8.86. The mean neurotic score of the deaf girls, 35.90, is also much greater than the 24.22 mean of the hearing girls. The difference of 11.68 is statistically significant, for the critical ratio is 5.53. When the sexes are combined, the mean neurotic score of the total deaf children, 36.75, is more than one and a half times as large as that of the hearing children. The mean difference of 14.35 has a very high degree of statistical reliability, for the critical ratio is 10.32.

The standard deviations for the various distributions of deaf and hearing children are very large. In all instances, the standard deviations are almost one-half as large as the mean neurotic scores for their distributions. Since the middle sixty-seven per cent of the cases of a distribution fall between ± 1 standard deviation, the results indicate that there is a wide range of neurotic scores. The distribution of scores for the deaf and hearing children are skewed, since the Brown measures atypical symptoms. The scores range from very few neurotic responses, representing good emotional adjustment, to many neurotic responses, representing the presence of a great deal of emotional maladjustment. There is some overlapping of the individual scores between the deaf and hearing groups. About twenty-six per cent of the total deaf group receive scores that are at or below the mean for the total hearing groups, while eighteen per cent of the latter receive scores that are at or above the total deaf groups' mean. The sex differences are small and unreliable. The mean difference between the deaf boys and girls is 1.53, and the mean difference between the hearing boys and girls is 3.19. The critical ratios are .83 and 1.51 for the deaf and hearing respectively. This is in accord with Brown's findings.

When the results of this study are compared with the published decile norms for the Brown Personality Inventory,[1] the means for the deaf boys, girls, and combined sexes correspond to the tenth decile. The hearing boys correspond to the seventh decile, the hearing girls to the eighth decile, and the combined hearing boys and girls correspond to the seventh decile. These comparisons reveal that the deaf and hearing children included in this study receive much higher neurotic scores than the children of the norm group. The deaf children in particular fall within the highest possible range, the tenth decile

[1] Brown, F.: *Suggestions with Regard to Use and Interpretations of the Brown Personality Inventory.* New York: Psychological Corp., 1935, p. 2.

and, according to Brown's classification, they make a "*very poor adjustment.*" The hearing groups fall within the "*poor adjustment*" category.

2. INFLUENCE OF CHRONOLOGICAL AGE AND INTELLIGENCE

The relationships of the neurotic score on the Brown inventory to chronological age and intelligence were determined by the correlation technique. The results are tabulated in Table III. The correlations

TABLE III.—CORRELATIONS OF THE BROWN NEUROTIC SCORE WITH CHRONOLOGICAL AGE AND INTELLIGENCE

Measures correlated	Group	Deaf			Hearing		
		N	*r*	PE_r	*N*	*r*	PE_r
Brown score with chronological age	Boys	225	.341	.040	190	.045	.048
	Girls	172	.166	.050	137	.392	.049
	Total	397	.270	.031	327	.133	.036
Brown score with Goodenough IQ	Boys	209	.095	.046	168	.041	.052
	Girls	167	.038	.052	127	.011	.060
	Total	376	.096	.035	295	.027	.039

between the neurotic scores and chronological age range from .045 ± .048 to .392 ± .049. The highest correlations, .341 and .392, are for the deaf boys and hearing girls respectively. For the total deaf group, the correlations is .270 ± .031, and for the total hearing group, the correlation is .133 ± .036. Both of these relationships are very low, but they are higher than the one reported by Brown[1] of .009 ± .03, between neurotic scores and age. The correlation for the total deaf group, while low, is nevertheless reliable, since it is almost nine times its probable error. While Brown's results are based on a younger group, it is doubtful whether chronological age was a big factor in the high neurotic scores obtained by the children in this study. Chronological age may have had a slight effect on the results, but the influence of age is small. There was a tendency for the older deaf boys and hearing girls to get higher neurotic scores. But the chronological age has no reliable relationship with the deaf girls' and hearing boys' neurotic scores.

[1] Brown, F.: A Psychoneurotic Inventory for Children Between Nine and Fourteen Years of Age. *Journal of Applied Psychology*, Vol. XVIII, 1934, p. 575.

Earlier in this article, it was reported that there was no reliable difference between the deaf and hearing boys on the Goodenough IQ. The hearing girls were superior to the deaf girls and the difference between these groups was statistically significant. The correlations between the neurotic score on the Brown inventory and intelligence, (Table III) range from .011 ± .060 to .095 ± .046. All of these are very low and are highly unreliable. These results indicate that there is no relationship between the neurotic score on the Brown inventory and intelligence. They are in accord with Brown's finding of a correlation of .013 ± .03 with IQ.

3. INFLUENCE OF AGE AT ONSET OF DEAFNESS

Information in regard to the age of onset of deafness was available for two hundred fifteen deaf boys and one hundred sixty-two deaf girls. The contingency correlations between this variable and the Brown neurotic score are .13 for deaf boys and .39 for deaf girls. These contingency correlations indicate that the deaf boys' neurotic score is not influenced by the age at onset of deafness. There is a tendency for the girls who become deaf at a later age to get higher neurotic scores. These correlations are not very high, and they must be interpreted with caution.

4. INFLUENCE OF THE PER CENT OF HEARING LOSS

Product-moment correlations between the Brown neurotic scores and the 3*A* audiometer hearing loss scores were computed for two hundred five deaf boys and one hundred sixty-nine deaf girls. The correlations are −.021 ± .047 and −.081 ± .055 for deaf boys and girls, respectively. These results indicate that the per cent of hearing loss in the better ear has no influence on the neurotic scores of the deaf children. The correlations are very low and unreliable.

5. SIGNIFICANT ITEMS ON THE BROWN PERSONALITY INVENTORY

The individual responses on the items of the Brown Personality Inventory were studied for the total group of deaf and hearing children. The percentage of symptomatic responses on each item in the inventory was calculated. The significance of the percentage differences for the deaf and hearing was then calculated in terms of the standard error of the difference. The results indicate that there is a tendency for the deaf children to give more symptomatic responses than the hearing children on practically all the questions of the inventory. In sixty-

nine out of the eighty questions, the percentage of symptomatic responses of the deaf children is larger and statistically significant, *i.e.*, the critical ratios are above 3.00. On only one question, "Have you been told at home that children should be seen and not heard?" is the percentage of symptomatic response of the hearing children larger and statistically significant.

6. RELIABILITY OF THE BROWN PERSONALITY INVENTORY RESULTS

In order to determine whether the results on the Brown Personality Inventory were equally reliable for both the deaf and hearing children, the coefficients of correlation between the halves of the inventory were computed for both groups. Uncorrected reliability coefficients, ranging from .902 to .943, are found for both the deaf and hearing groups. When the Spearman-Brown prophecy formula for the whole inventory is applied to these results, the corrected correlations range from .948 to .971. These reliability coefficients are even higher than those reported by Brown. The latter reports corrected reliability coefficients ranging from .860 to .928. The above results indicate that the Brown neurotic results for both the deaf and hearing children possess a very high degree of reliability.

SUMMARY AND CONCLUSIONS

The Brown Personality Inventory for Children was given to three hundred ninety-seven deaf and three hundred twenty-seven hearing children, in order to determine whether there are group differences between deaf and hearing children in their manifestation of maladjusted behavior, as measured by their psychoneurotic responses. Difficulty was encountered in securing an adequate hearing control group of the same chronological age as the deaf group. Because of their verbal handicap, it was impossible to give this inventory to the younger deaf children. The deaf children were, on the average, more than four years older than their hearing control. These deaf and hearing children were comparably equated on nationality and general social status. The deaf girls were inferior to the hearing girls on intelligence, while no difference existed between the deaf and hearing boys on this variable.

All the groups of deaf children were found to receive much higher neurotic scores than the hearing control children. The differences between the deaf and hearing children were very large and of a high

degree of statistical significance. No reliable differences were found between the sexes of the deaf and hearing groups.

When the results of this study were compared with the published norms, all the deaf and hearing groups were found to receive much higher neurotic scores than the children of the norm group. The deaf children, in particular, fell within the highest possible range and, according to Brown's classification, made a "*very poor* adjustment." The hearing groups fell within the "*poor adjustment*" category.

The relationship between chronological age and the Brown neurotic scores was found to be .270 ± .031 for the total deaf group, while it was .133 ± .036 for the total hearing group. There was a slight tendency, on the whole, for the older deaf boys and the older hearing girls to receive slightly higher neurotic scores.

The relationship between the neurotic scores and intelligence was found to be very low and insignificant. It varied from .011 ± .060 to .095 ± .046, for the deaf and hearing groups.

There was a tendency for the girls who became deaf at an older age to have higher neurotic scores. This conclusion was borne out by a contingency correlation of .39 for deaf girls. The neurotic scores of the deaf boys were uninfluenced by the age at onset of deafness ($c = .14$). The neurotic scores were not influenced by the amount of residual hearing that the deaf children possessed. Low, insignificant correlations of −.021 ± .047 and −.081 ± .055 were found for the deaf boys and girls, respectively.

The deaf children, when compared with their hearing control, showed reliable differences on sixty-nine items of the Brown Personality Inventory. On only one item was the percentage of symptomatic response of the hearing children statistically significant.

Reliability coefficients between the halves of the Brown Personality Inventory were found to be equally high for the deaf and hearing groups. The halves of the inventory yielded reliability coefficients ranging from .902 to .943. When the Spearman-Brown prophecy formula was applied, these correlations were found to range from .948 to .971.

A PRELIMINARY STUDY OF RELATIONS BETWEEN ATTITUDE, ADJUSTMENT, AND VOCATIONAL INTEREST TESTS[1]

JOHN G. DARLEY

The General College, University of Minnesota

In the simultaneous clinical use of several kinds of tests, certain inter-relations are empirically identified in student interviews. Often these relations require the clinical readjustment of massed statistics for the individual case, as, for example, in marked discrepancies between ability and achievement. The diagnosis of under-achievement leads in turn to clinical treatment of the underlying causes of the discrepancy. In some instances empirically observed relations tend to give meaning to a test recently introduced into a testing program in terms of the new test's relation to a test whose utility or meaning is already known to the counselor. Thus, a new test of vocational interests or of personality characteristics or of achievement may be shown empirically and statistically to have a relation to existing tests in the same behavior areas. In still other instances, relations between different kinds of tests may serve to round out, complement, or extend the counselor's stock of occupational ability profiles or professional training group profiles.[2] Quite often these empiric relations are "hunches" on the part of the counselor whereby he judges the student to have potentially many of the characteristics necessary for certain kinds of job competition; or they are "hunches" regarding an apparently elusive connection between differing kinds of tests. Verifications of such clinical clues are valuable additions to the counselor's fund of descriptions of workers or potential workers.

These profiles of families of occupations or of professional training groups are the functional basis of much educational and vocational counseling and may be briefly described. Members of various occupational or training groups may be differentiated in terms of abstract ability, specific aptitudes, measured occupational interests, specific achievement or proficiency measures, and personality characteristics.

[1] The author wishes to acknowledge assistance received in this study under Works Progress Administration project number 6094-4214.

[2] For the origin and extension of this concept of occupational ability profiles, see Trabue, M. R.: "Occupational Ability Patterns." *The Personnel Journal*, Vol. II, 1933, pp. 344–351; and Dodge, A. F.: *Occupational Ability Patterns*. Bureau of Publications, Teachers College, Columbia University, 1935.

These unit analyses of various behavior areas are then synthesized into profiles that differentiate one group from another and from people in general. As more areas of behavior fall within the scope of measurement, additional group differences may be found to exist. Occupational ability profiles assume also the identification of groups performing the same job or training functions, regardless of job or training label, and the application to such groups of a uniform program of test analysis.[1]

It is immediately granted that such profiles do not cover *all* characteristics basic to job or training success, but it is equally obvious that the concept permits greater accuracy in matching the individual to the group within which he is most likely to compete satisfactorily. Since such matching is basic to a valid program of educational-vocational guidance, the occupational ability profile offers at once a concept and a method for more effective guidance.

Data for the present study were drawn from the administration of a four-to-five-hour testing program to students in the General College of the University of Minnesota. This program included measures of abilities, achievement, aptitudes, vocational interests, attitudes, and personality adjustments. Selected test scores for samples of men were analyzed statistically on the basis of the relations empirically observed in interviews between measured vocational interests and measured attitudes or adjustments. It was felt that such analyses might serve as further test definitions of groups with differentiated occupational interest patterns.

The Strong Vocational Interest Test for men,[2] the Minnesota Scale for the Survey of Opinions,[3] the Bell Adjustment Inventory,[4] and the Minnesota Inventories of Social Attitudes[5] were studied. The

---

[1] See Dvorak, B. J.: *Differential Occupational Ability Patterns.* Bulletins of the Employment Stabilization Research Institute, Vol. III, No. 8, University of Minnesota Press, Minneapolis, February, 1935.

[2] Strong, E. K. Jr.: *Manual for the Vocational Interest Test.* Stanford University Press, 1936.

[3] Rundquist, E. A., and Sletto, R. F.: *Personality in the Depression.* University of Minnesota Press, 1937; and *Manual for the Minnesota Scale for the Survey of Opinions.* University of Minnesota Press, 1937.

[4] Bell, H. M.: *The Theory and Practice of Student Counseling.* Stanford University Press, 1935; and *Manual for the Adjustment Inventory.* Stanford University Press, 1935.

[5] Williamson, E. G., and Darley, J. G.: *Manual for the Minnesota Inventories of Social Attitudes.* The Psychological Corporation, New York City, 1937.

Minnesota Scale for the Survey of Opinions is scored for morale, feelings of inferiority, family adjustments, attitudes toward law, economic conservatism, and attitudes toward education. The Adjustment Inventory is scored for home, health, social, and emotional adjustments. The Minnesota Inventories of Social Attitudes are scored for social preferences and social behavior.

The Strong Vocational Interest Test for men is scored for the following occupations:

I. Technical—Chemist. Engineer. Scientific Farmer. Mathematician. Physicist. Architect. Dentist. Physician. Psychologist. Artist.
II. Verbal—Lawyer. Editor. Advertising Agency Man.
III. Business Contact—Life Insurance Salesman. Real Estate Salesman.
IV. Social Service *A*—Minister. Teacher.
V. Social Service *B*—Personnel Manager. City School Superintendent. Y. M. C. A. General Secretary. Y. M. C. A. Physical Director. Y. M. C. A. Boys' Secretary.
VI. Business Detail—Accountant. Office Clerk. Purchasing Agent. Specialty Salesman.
VII. Certified Public Accountant.

The labels applied to the groups of occupations are adaptations of the descriptive titles used in the factor analysis presented by Strong.[1] They serve to identify basic interest patterns, and in clinical use of the test, it has appeared, locally at least, that best results are obtained in most cases by emphasizing the individual's primary interest pattern, and considering specific keys as typical occupations within the pattern rather than as perfect occupational niches for the individual. This pattern analysis permits greater latitude to the counselor in harmonizing interests, abilities, achievement, and training or job opportunities.

In the course of a detailed case analysis at the end of the 1935–1936 school year, two trained counselors made clinical judgments of the primary, secondary, and tertiary interest patterns for each man who had taken the Vocational Interest Test. Judgments of the absence of such patterns were also made. Judgments of primary patterns were based on a preponderance of *A* and *B*+ scores on the specific keys within any grouping; judgments of secondary patterns were based on a preponderance of *B*+ and *B* scores on the specific keys within any grouping; judgments of tertiary patterns were based on a preponderance of *B* and *B*− scores on the specific keys within any grouping.

[1] Strong, E. K., Jr.: "The Vocational Interest Test." *Occupations*, Vol. XII, No. 8, April, 1934, pp. 49–56.

Thus it would be possible for an individual to have a primary pattern of "technical" interests, a secondary pattern of "business detail" interests, and a tertiary pattern of "social service *A*" interests. It would also be possible for an individual to have no primary or secondary patterns, with a tertiary pattern of "verbal" interests. Many other such combinations naturally occurred: One primary pattern, no secondary or tertiary; a primary and secondary, no tertiary; no primary, one secondary, no tertiary. This method of analysis did not do too great violence to the scoring structure of the test, and actually seemed to enhance its clinical utility.

In the present study, only the existing primary interest patterns have been used. Mean scores and sigmas on the attitude and adjustment tests of groups of students having different primary interest patterns were used in deriving the critical ratios appearing in Table I. The following listing gives first the labels of the basic interest groupings, followed by the number of students having primary patterns in the group:

	Cases
Technical	30
Verbal	34
Business Contact	33
Social Service *A*	18
Social Service *B*	35
Business Detail	103
Certified Public Accountant	0

There were in addition one hundred thirty men for whom no primary interest group was judged to exist. These cases do not appear in Table I.

Table I summarizes all critical ratios for cases having the six primary patterns as indicated. Men with primary patterns in each grouping have been compared with men having primary patterns in every other grouping in respect to mean differences on the twelve attitude and adjustment measures indicated at the top of the table. The small number of cases showing various primary interest patterns may act to cut down the critical ratios presented; for this preliminary analysis, therefore, a critical ratio of 2.00, indicating ninety-eight chances out of one hundred for the recurrence of mean differences in the same direction, may be considered a tentative critical limit.

Men with primary interest patterns in technical occupations achieve poorer morale scores than men in all other groups but verbal

Table I.—Summary of Critical Ratios between Mean Attitude and Adjustment Scores Made by Men with

TABLE I.—SUMMARY OF CRITICAL RATIOS BETWEEN MEAN ATTITUDE AND ADJUSTMENT SCORES MADE BY MEN WITH DIFFERENT PRIMARY INTEREST PATTERNS

	Minnesota scale for survey of opinions						Adjustment inventory				Minnesota inventory of social attitudes		*N*
	Morale	Inferiority	Family	Law	Economic conservatism	Education	Home	Health	Social	Emotional	Social preferences	Social behavior	
Technical *vs.* Verbal	0.86	0.63	−0.66	1.64	1.54	1.09	−2.03	−1.09	3.05	−2.32	1.02	2.07	30 *vs.* 34
Technical *vs.* Business Contact	2.05	2.03	0.99	1.57	1.58	0.72	−1.63	−0.94	4.54	−1.53	3.56	3.49	30 *vs.* 33
Technical *vs.* Social Service *A*	1.97	1.06	0.08	1.36	0.79	0.05	−0.78	−0.79	2.46	−1.19	2.26	1.24	30 *vs.* 18
Technical *vs.* Social Service *B*	2.83	3.24	1.67	2.31	0.54	0.73	−1.41	−0.38	4.44	−0.25	4.38	2.60	30 *vs.* 35
Technical *vs.* Business Detail	2.13	1.10	0.68	1.38	2.76	0.90	−0.79	−0.53	2.52	−0.53	3.23	1.56	30 *vs.* 103
Verbal *vs.* Business Contact	1.14	1.40	1.62	−0.08	0.85	−0.19	0.41	0.16	1.68	0.62	3.15	1.34	34 *vs.* 33
Verbal *vs.* Social Service *A*	0.87	0.40	0.78	−0.51	−0.76	−0.95	1.14	0.25	0.11	0.67	1.51	−0.95	34 *vs.* 18
Verbal *vs.* Social Service *B*	1.74	2.45	2.32	0.49	−0.35	−0.36	0.73	0.78	1.74	1.48	4.32	1.06	34 *vs.* 35
Verbal *vs.* Business Detail	0.97	0.27	1.47	−0.57	0.41	−0.43	1.64	0.82	−1.00	2.23	1.38	−0.91	34 *vs.* 103
Business Contact *vs.* Social Service *A*	−0.50	−1.09	−0.97	−0.43	−1.67	−0.63	0.76	0.09	−1.08	0.12	−2.04	−2.50	33 *vs.* 18
Business Contact *vs.* Social Service *B*	0.44	0.78	0.53	0.68	−2.30	−0.11	0.31	0.62	0.20	0.82	0.35	0.02	33 *vs.* 35
Business Contact *vs.* Business Detail	−0.50	−1.47	−0.57	−0.58	−0.68	−0.12	1.18	0.65	−3.00	1.31	−1.78	−2.75	33 *vs.* 103
Social Service *A* *vs.* Social Service *B*	1.20	2.15	1.69	1.17	−0.38	0.62	−0.50	0.48	1.18	0.60	3.13	1.77	18 *vs.* 35
Social Service *A* *vs.* Business Detail	0.04	−0.22	0.64	−0.03	1.97	0.73	0.22	0.50	−0.78	0.95	0.62	0.20	18 *vs.* 103
Social Service *B* *vs.* Business Detail	−1.30	−2.97	−1.44	1.27	2.52	0.00	0.91	−0.06	−2.88	0.29	−3.06	−1.78	35 *vs.* 103

N.B.—No sign before a critical ratio indicates a tendency toward poorer measured adjustments and/or greater radicalism in the first group of each pair. A minus sign before a critical ratio indicates a tendency toward poorer measured adjustments and/or greater radicalism in the second group of each pair.

occupations. These men are characterized also by somewhat greater feelings of inferiority than men with primary interests in business contact or social contact jobs (Social Service *B*). They also seem to have markedly poorer social adjustments, fewer social preferences, and more limited social skills than men in nearly all other primary interest groupings. On the favorable side, they show a tendency for consistently better home, health, and emotional adjustments (all critical ratios in these three columns are prefaced by the minus sign indicating poorer measurement results for the second group in each pair). They show a similar tendency for less conservatism in regard to legal, economic, and educational questions than is found in men in the other primary interest groupings.

Some aspects of the foregoing sketch correspond to existing stereotypes of technical students on the campuses of the country—disinterested if not intolerant in regard to the more refined social amenities, and somewhat inept in social activities. Insofar as the occupations within this pattern also include some academic workers, the tendency to limited socialisation, greater liberalism, and less dominance or self-assuredness is again typical. Finally, the evidence on the earlier maturation of this interest pattern may be related in some way to the attitude and adjustment characteristics of these men with primary interests in technical activities.

Men with primary interest patterns in occupations of the Verbal or Social Service *A* groupings do not show characteristic attitudes or adjustments that set them significantly apart from men with other primary patterns. However, they *tend* to have poorer morale, greater feelings of inferiority, poorer home or family, social, and emotional adjustments, more limited social preferences and less sure-footedness in social situations than men with primary patterns in the business contact and Social Service *B* occupations. It is also to be remarked that the occupations grouped as Verbal and Social Service *A* tend to fall at the less masculine end of Strong's scale designed to measure masculinity of interests. The relation between this latter finding and the sketch of the attitude and adjustment attributes of these men is as yet obscure.

Men with primary interest patterns in Business Contact jobs, or sales jobs, are clearly better in social adjustments, have greater social preferences, and are more socially adept than men in all other groups except Social Service *B*, social contact jobs. These men with sales interests are, however, equally clearly differentiated from men in the

Social Service *B* group by their greater economic conservatism. Note also the marked difference in social adjustments, preferences and skills between the Business Contact and the Business Detail workers; these differences lend considerable support to stereotypes of the aggressive salesman and the meek bookkeeper. Salesman and bookkeeper alike, however, are closely similar in regard to legal, economic, and educational conservatism.

These sketches summarize the outstanding trends in Table I. They offer interesting extensions of occupational ability profiles into the realm of personality characteristics that would seem to be developmentally determined in the individuals prior to occupational training or competition. These attitude and adjustment tests not only derive meaning from their relations to the vocational interest test, but also add meaning to it by complementing its definitions of occupational interest types.

This preliminary study suggests at least three further lines of research. More intensive treatment of paired attitude or adjustment and interest test scores would permit the determination of relations obtaining for individuals rather than groups. A study of test-retest changes in attitudes and adjustments within primary interest groups would throw some light on factors influencing differential changes in attitudes or adjustments. As a corollary to this longitudinal study, a cross-section attitude and adjustment testing of employed adults known to have primary interest patterns in the various occupational groupings would indicate the presence or absence of these trends in men actually engaged in job competition.

Meanwhile, the present data should be immediately useful to counselors in educational and vocational guidance. Ruml's concept of the "worker-in-his-work" unit may come to include not only a balance between interests, opportunities, and abilities, but also between attitudes and adjustments. Trabue's occupational ability profiles may be based not alone on abilities, aptitudes, proficiencies, and interests, but may include characteristic attitudes and adjustments.

BOOK REVIEWS

WILLIAM STERN (tr. H. D. Spoerl). *General Psychology from the Personalistic Standpoint.* New York: Macmillan, 1938, pp. XXII + 589.

Here is another brand new system of psychology. Well, not quite, for it has many likenesses to organismic psychology as it is expounded on this continent. Like gestalt psychology, it comes from the fertile brain of a German. While not quite as closely argued as Koffka's *Principles of Gestalt Psychology*, it is even wider in scope, taking the whole field of psychology as its province.

Stern defines psychology as "the science of the person having experience or capable of having experience." In six long parts, he expounds his system and fits the results of an enormous number of experimental findings, mostly from German laboratories, within his scheme. In this he shows new aspects of old material and gives a personalistic interpretation of them. Some of his explanations will be questioned by most readers. In a section entitled "acquired reflexes" which deals with conditioned reactions he states: "The chauffeur who has learned that one must pass approaching obstacles on the right, ends by making the proper twist of the steering wheel from any obstacle that looms up suddenly before him. Normally his action is purely reflex."

The book is well-bound, but the type is so small that continued reading tired my eyes. I must confess that I found it rather stodgy, despite an obviously fine translation, but it may be that reading such a work at the end of a long session was not quite fair to it. Probably a second reading would lead to a more favorable opinion. I think it is a good but not an epoch-making book. PETER SANDIFORD.

University of Toronto.

JEAN MARQUIS DEUTSCHE. *The Development of Children's Concepts of Causal Relations.* University of Minnesota Press, 1937, pp. 104.

This monograph reports an outstanding experimental approach to the basic problem of the nature of the thought processes in large numbers of children of school age, and as such it represents a valuable and unique contribution to the field of child psychology. Piaget's classification of children's concepts of causal relations was used as the

point of departure and seven hundred thirty-two children ranging in age from eight to sixteen years were asked to give written explanations of a series of eleven natural phenomena, each of which was demonstrated in the classroom. A second form of the test involved the spontaneous written answers to twelve questions such as "What makes the wind blow?" and "What makes shadows?" Thirteen raters who were supplied with physicists' explanations of the phenomena in question rated the answers on an arbitrary eight-point scale, each step of which was carefully defined. Their ratings proved to have very satisfactory reliability. This quantitative approach revealed regular age and grade trends indicating increasing accuracy of causal explanations, with boys maintaining a slight and consistent superiority over girls. No consistent relationship to socio-economic status was found.

A qualitative analysis was conducted involving the attempt by three raters to classify the answers into Piaget's seventeen types of causal thought. Since three trained raters disagreed considerably and numerous other difficulties were encountered in this analysis, the author was led to three basic criticisms of Piaget's classification: "(1) That the answers actually given do not fit naturally into these types; (2) that the types are of unequal value, some being too inclusive and some so specific as not to be represented at all by the data; and (3) that personal judgment enters to a great extent into the classification." The most striking result of this analysis was that only four of Piaget's seventeen types of causal thinking (phenomenistic, dynamic, mechanical and logical deduction) were found in large enough frequencies to warrant age analyses. Phenomenistic explanations were found to decrease with age, whereas mechanical explanations and logical deductions increased with age. A high degree of specificity was observable in the answers, indicating that the type of question rather than the level of the child's thought processes is of primary importance in determining the type of response. Even at the eight-year level eighty-one per cent of the answers were materialistic, indicating that the animistic, dynamic, and prelogical types, claimed by Piaget, must have disappeared by this age, if they ever existed.

The author concludes that "No evidence was found that children's reasoning develops by stages [as Piaget claims]. Both quantified and qualitative analyses show a gradual progression in answers with advancing age, and all kinds of answers are found spread widely over the age range." The author further found that in general these

children were much more realistic and scientific in their explanations, and were less inclined to give magical and fantastic explanations than were the children studied by Piaget. She states that "All investigators who have repeated Piaget's questions find a much higher percentage of naturalistic answers than Piaget reports, and find that prelogical answers such as Piaget obtains in high frequency are made conspicuous by their rarity."

Thus is added to the American child psychology literature one more well-executed experimental study, which was stimulated by Piaget's work, but which resulted in conclusions directly opposed to his claims.

DOROTHEA MCCARTY.

Fordham University.

DANIEL A. PRESCOTT. *Emotion and the Educative Process.* Washington: American Council on Education, 1938, pp. xviii + 323.

This is a report to the American Council on Education by its Committee on the Relation of Emotion to the Educative Process. The first half of the book is devoted to a generalized discussion of affective responses as revealed in the literature. Evidence on physiological basis, patterning, trainability, and maturity of affective experience are analyzed. Also, the relation to social institutions and to learning are considered together with personality needs and conditions which frustrate them. In the latter half of the treatise the implications of affection for education are discussed.

It is apparent from the survey that little is known concerning the rôle of affective factors in personality and in education. Although there have been many studies of feelings and emotions, few have attempted to consider affective factors as directly related to educational practice. The principal findings of this survey, therefore, have been to reveal ignorance on the subject and to point out the numerous problems that are awaiting solution. There are, of course, certain results which have educational significance. At present, however, these can only be stated in general terms. Knowledge of how they may be applied directly to the educative process awaits further investigation. Various areas deserving of extensive research are outlined. Because they are complex and comprehensive, it is suggested that coöperative effort be employed in attacking the problems.

The systematic study of feelings and emotions has been far from satisfactory. Although there may be general agreement on the importance of affective experience as a factor in education, the lack of

adequate techniques of investigation probably means that progress in this field will be relatively slow.

Although some will not agree with certain of the inferences derived from the survey, it must be admitted that the Committee has done a worth-while task in a commendable manner. MILES A. TINKER.

University of Minnesota.

HENRY DANIEL RINSLAND. *Constructing Tests and Grading in Elementary and High School Subjects.* New York: Prentice-Hall, Inc., 1937, pp. IX + 323.

The first two thirds of this text are quite similar to Part II of Ruch's *The Objective or New Type Examination* (1929). After a brief introductory chapter in which the unreliability of essay tests and teachers' marks in general are stressed, the author devotes the next two hundred pages to a consideration of the various types of objective questions. The multiple choice item, for example, is first defined, then sixteen rules for its construction are stated and illustrated with a "poor" and "good" question under each rule, after which come thirty pages of "Samples of Multiple Choice Examinations in Elementary and High School Subjects." A similar treatment is afforded the sentence completion, simple recall, matching, and alternative response types of objective questions.

The suggestions for construction in each instance are numerous and appropriate. The illustrative examinations include many ingenious and interesting teacher-made tests. To the reviewer the most thought provoking section of Rinsland's book was Chapter VII, "Miscellaneous Forms (of Objective Questions)" (pp. 143–223). These miscellaneous forms illustrate well the originality of many classroom teachers in adapting the objective testing technique to their own needs and purposes. A number of these examinations indicate that some teachers have discriminated between tests that measure things memorized and tests which determine whether pupils can go beyond the reproduction of facts and demonstrate ability to use these facts in solving problems.

Chapters VIII and IX are entitled, respectively, "The Improvement of Marks and Grading Systems" and "General Structure and Value of Objective Tests." The former recommends that absolute marks be discarded in favor of relative marks determined by application of the normal curve concept. Chapter IX is a discussion of the

objective test in terms of its reliability and validity. Methods are presented for the determination of each. The text ends with a tabulation of the merits and demerits of the objective form of question.

Rinsland, in common with many persons who write books on tests, implies that the objective form of question is the only form that enables a teacher to measure. He says very little, other than by way of destructive criticism, about the essay examination. There is no extensive discussion of the value of essay tests, nor, with the exception of a short section on "Rating Subjective Examinations" (pp. 252–257), are there any profound suggestions for the improvement of this type of evaluational instrument. The reviewer is of the opinion, and this opinion is not supported by subjective evidence only, that the essay test not only has value but that this value can be greatly enhanced by instruction in the technique of essay test construction and scoring.

Rinsland, too, does not discuss adequately the nature of the mental processes which are involved in arriving at the correct answer to various types of objective questions. With the exception of some of the miscellaneous examinations, most of the illustrative tests represent attempts to measure memory and fixed associations only. The possibility of using objective questions to measure reasoning, the application of principles, the ability to infer, and other higher mental processes was not developed sufficiently in view of the importance of the topic. Throughout the text, for example, the reviewer could find no discussion of the important work which is being done on the technique of testing by the evaluation specialists of the Progressive Education Association.

STEPHEN M. COREY.

University of Wisconsin.

DAVID H. RUSSELL. *Characteristics of Good and Poor Spellers: A Diagnostic Study.* New York: Bureau of Publications, Teachers College, Columbia University, 1937, pp. 103.

This monograph is number 727 in the invaluable Teachers College Contributions to Education. The author has set the stage for his own investigation by a careful analysis of preceding work in the teaching of spelling and the study of spelling disability.

The original contribution of the monograph is based upon an exhaustive study of sixty-nine pairs of pupils from the third, fourth, and fifth grades carefully matched on the basis of school, sex, grade, chronological age, IQ, and terms in school. One member of each pair

was one year or more retarded in spelling. To these sixty-nine pairs of children an elaborate battery of tests, including handwriting, diagnostic spelling and reading, vision, hearing, speech, and silent reading was given. This battery was analysed into eighty-three different elements on which the two groups were compared. The statistical data concerning the comparison on each of these elements is given in full, but for the purpose of review we can only summarize the conclusions.

The etiology of spelling retardation is unique with the individual child. In general, however, organic factors of vision and hearing were of minor importance. Next most important were methods of studying spelling, about which Russell comments that "systems of studying spelling are urgently needed which are less formal and ritualistic." The area in which the greatest number of reliable differences were found was academic achievement in reading, handwriting, and spelling. This finding does not indicate that reading and handwriting are basic to spelling. Rather it suggests that there are a group of basic word skills necessary for success in any language activity. If these are not acquired early in the school career difficulty may appear in any or all language activities.

It appears to the reviewer that this investigation is a major contribution to the understanding of spelling difficulties in school children. The author has presented some pointed comments on teaching methods in spelling, and on the correction of difficulties in this field.

C. M. Louttit.

Indiana University.

M. D. Vernon. *Visual Perception.* Cambridge, England: Cambridge University Press, 1937, pp. xi + 247.

This is the first comprehensive account in English of the important experimental work on visual perception. Some major aspects considered include development of the perceptual process, the parts played by the observer and by the stimulating environment on the percept, contributions of Gestaltists to studies on visual perception, and perceptual types. A description of the important tachistoscopes is appended. No marked theoretical bias occurs in interpretation. The author endeavors to demonstrate that "perception is an intricate process of structural differentiation from among innumerable excitations presented by the physical environment." It would appear that

the primary state of vague awareness becomes differentiated into a detailed and comprehensible perceptual pattern whose form is determined by individual configural tendencies, past experience, innate tendencies, attitudes, physiological condition, and by restrictions of the environment (in certain laboratory set-ups). The phenomenal percept is largely determined, therefore, by established functional schemata.

The author has maintained a critical attitude throughout the discussion. Her attempt at coördination of materials has been fairly successful, especially when one considers the complexity of this field. Some readers, however, will note certain shortcomings of the book. Prominent among these are too brief and incomplete treatment of eidetic imagery, illusions, and upright vision. The list of three hundred seventy-seven references is extensive but presented in an inadequate and annoying manner. The title of each reference is omitted and the page numbers are incomplete. On the other hand there are certain topics, such as figure and ground relations and synaesthesia, whose treatment is highly commendable.

The author should be complimented on the satisfactory completion of a difficult task. No psychologist interested in perception can afford to be without this treatise.

MILES A. TINKER.

University of Minnesota.